AD_____

Award Number: W81XWH-04-1-0323

TITLE: Task-Specific Optimization of Mammographic Systems

PRINCIPAL INVESTIGATOR: Robert Saunders, Ph.D.

CONTRACTING ORGANIZATION: Duke University
                                            Durham NC 27710

REPORT DATE: March 2007

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                        Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-03-2007	Annual Summary	15 Feb 04 – 14 Feb 07

4. TITLE AND SUBTITLE

Task-Specific Optimization of Mammographic Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0323

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Robert Saunders, Ph.D.

E-Mail: rss@duke.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University
Durham NC 27710

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT:

This study sought to understand how different imaging parameters affect clinical diagnosis. First, it developed research tools for measurement and simulation of mammographic imaging. Second, we applied these research tools and conducted a large human observer experiment to answer several clinically relevant questions. The first question explored the impact of display resolution on the detection of breast masses and calcifications. We found that different displays had little impact on clinical performance. The second question explored the effect of reduced dose on the detection of breast lesions. We found that the increased noise from reduced dose did impact radiologist performance. Reducing the dose by half did not have a statistically significant impact on diagnostic accuracy, suggesting that mammographic dose could be reduced modestly with little impact on clinical performance. These results have immediate implications for clinical breast imaging.

15. SUBJECT TERMS
X-ray Imaging, Digital Imaging

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	171	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

The primary screening tool for breast cancer is x-ray mammography. While mammography reduces breast cancer mortality, it has areas for improvement as it misses many early-stage cancers. This research seeks to improve the efficacy of mammography by optimizing the entire image chain for the detection of breast masses and microcalcifications. This research can be split into two stages. The first stage measures the imaging chain's physical characteristics. These characteristics include resolution and noise measurements of x-ray detectors and medical displays. To better understand this physics, this research also has developed models of scattered radiation, as scatter is another major factor affecting resolution and noise. This physical data is then applied in the second research stage. The second stage modifies the resolution and noise of mammographic images. These images are viewed by a combination of observer models and human observers to discover how image quality affects lesion detection and discrimination. This observer data will help guide future optimization of mammographic systems.

Body

This section summarizes the research accomplished over the course of the entire grant period. We have written the sections of the statement of work and then noted our research accomplishments in each area.

Task 1: Create a simulation procedure for the anatomical background of mammographic images

1.1 *Acquire normal mammograms obtained on digital systems for analysis*
Working with colleagues from Emory University, we obtained 984 images acquired on an indirect flat-panel detector.

1.2 *Categorize the images into the four types of breast composition, as identified by the BIRADS system.*
Using the semi-automated technique proposed by Sivaramakrishna, *et al* in 2001,[1] we analyzed each of the mammograms in the above image database. This analysis gave us the percent of the breast area covered by fibroglandular tissue. However, we were not confident that this analysis reproduced the radiologist's assessment of breast density and did not use these categories in further analysis.

1.3 *Analyze the geometrical features of these breasts and characterize them with a fixed number of scalar parameters, such as size.*
These two steps were included as they would aid in the creation on a routine to simulate mammographic backgrounds. As part of the research for anatomical simulation, we searched the literature for previous research on mammographic background simulation. We discovered and implemented the methods of Bochud, *et al* to emulate mammographic backgrounds by creating clustered lumpy backgrounds.[2] These simulated backgrounds appeared similar to real mammographic backgrounds, but did not capture all of the complexity of real anatomy. Therefore, we decided to use the mammographic data set obtained in 1.1 for our subsequent simulation experiments.

1.4   *Obtain mammograms from the Digital Database for Screening Mammography (DDSM) to analyze lesion characteristics*
We selected images from the DDSM that contained oval circumscribed, oval obscured, irregular ill-defined, and irregular spiculated masses. In addition, we selected images that contained fine linear branching and pleomorphic calcifications. We segmented these mammograms into regions of 2.56 cm x 2.56 cm centered on the mass or calcification.

1.5   *Analyze the features of specific lesion types*
We analyzed the images obtained from the DDSM to create a model of the radiographic appearance of breast lesions. This model was described in two publications (Appendices V and X).

1.6   *Create a program that can create images with breast anatomy and breast lesions that allows for user input of specific scalar parameters, such as size.*
We created a program that allowed one to insert simulated masses and calcifications into normal anatomical backgrounds. The details of this program were disclosed in two publications (Appendices V and X). In addition, this program served a crucial role in our recent research on the impact of display image quality and impact of radiation dose, as disclosed in Appendices I-IV.

As noted in 1.2 and 1.3, the simulated mammographic backgrounds lacked the complexity of real backgrounds. We therefore used actual digital mammograms for our simulation experiments.

1.7   *Establish mapping technique to determine grayscale values of image using sigmoid curve transformation.*
To conduct this task, an experienced mammographer reviewed the digital images obtained in 1.1 and window and leveled each mammogram to produce a clinically relevant appearance. We recorded the parameters for each image and fit a sigmoid curve to each window and level function. We then applied the appropriate transformation to each image in order to simulate the correct clinical appearance. This stage was disclosed in two publications (Appendices IV, IX).

Task 2: Calibrate a computational observer (observer model) to emulate the detection task performed by mammographers.

2.1   *Create a set of anatomical images with the four different background types and different lesions types using the above simulation routine.*
We had previously acquired a set of normal mammograms that contained images with each of the four different background types, ranging from extremely dense to almost entirely fat-replaced. Using this set of normal mammograms, we inserted simulated lesions using our lesion simulation routine[3-5] to create a large image set with three different types of lesions, benign masses, malignant masses, and malignant microcalcifications.

2.2   *Modify the resolution and noise of the images to that consistent with various digital systems.*
Using our verified noise modification routine,[6] we simulated the effects of imaging with reduced dose. We created images with noise characteristics emulating three dose levels—full clinical dose, half dose, and quarter dose. We altered the resolution of the images by displaying the images on three different medical

displays. These displays included an LCD, a normal CRT, and a CRT with degraded resolution.

2.3   *Perform an observer performance experiment with five mammographers.*
Five experienced mammographers viewed the image set on three displays using a custom graphical interface. This interface allowed the mammographers to rate the images as containing no lesion, a benign mass, a malignant mass, or microcalcifications. As shown in the table below, the mammographers viewed 2200 images, which included three resolutions levels (one for each display) and three noise levels corresponding to full, half, and quarter dose.

Table. Distribution of images in each resolution and noise category. The category of (Degraded CRT, Half Dose) was not evaluated in this experiment to reduce the duration of the human observer experiment.

	Full Dose	*Half Dose*	*Quarter Dose*
LCD	400	200	200
Normal CRT	400	200	200
Degraded CRT	400	--	200

2.4   *Analyze the data from that experiment with Receiver Operating Characteristic Analysis.*
Receiver Operating Characteristic Analysis significantly slows down an observer experiment because of the detailed ratings it requires. It also differs from the clinical paradigm by requiring radiologists to specify their confidence in a given decision. In the clinic, radiologists generally make binary decisions as to whether a lesion is present or not. Therefore, this experiment did not use Receiver Operating Characteristic Analysis, but rather used a new categorical rating paradigm that minimized reading time and more closely emulated clinical decision making.

We are analyzed the observer data to find overall classification accuracy at different dose levels and on different displays. As well, the data was analyzed for performance at specific clinical tasks, such as the detection of microcalcifications and discrimination of benign and malignant masses. Resolution and noise was considered separately and jointly to understand how these two parameters jointly affected lesion detection, discrimination, and decision-making time. Refer to the Appendix I-IV for four publications describing this analysis.

2.5   *Use several computational observers to examine the image set.*
We found that this image set was not appropriate for observer model calculations as it did not model resolution, but rather used displays for resolution modification. Therefore, we combined this step with specific aim 3.3, which analyzed images with different simulated noise and resolution characteristics.

2.6   *Using the observer model that best matches the performance of the mammographers, calibrate that model to the human performance.*
As described by the publication in Appendix I, we did not find any one observer model completely matched human performance at all detection and discrimination tasks. Observer model performance did not fully simulate human performance. Observer model results showed drops in detection and

discrimination with increased display blur, while human observers did not. Computational observers could be more sensitive to display blur than human observers and therefore further work must be conducted to optimize observer models. The NPWE model had difficulty detecting microcalcifications. This deficiency in the NPWE observer has also been noted in a previous study. [7] These points suggest that further work is still needed in optimizing observer models to fully replicate human performance.

Task 3: Create an empirical model that relates the resolution and noise of a digital mammographic system to the detectability of breast lesions.

3.1   Compile a list of MTFs and NPS for commercial radiographic systems, including image processing algorithms and displays.
We have conducted studies to directly measure the physical characteristics of mammographic systems. Please refer to two papers in Appendices VIII and XI for details about one study where we measured the performance of a clinical prototype digital mammographic system. We extended that work by conducting a study that measured the resolution and noise of five medical displays, disclosed in Appendix VI.

3.2   Create a set of 1500 simulated anatomical images with added masses and microcalcifications. The resolution and noise of these images will be modified according to the various configurations collected above.
An image set of 2200 images was created using similar methods as the one created under specific aim 2.1. In this case, an image set was created that had three different resolution levels, divided among images with resolution corresponding to an LCD, a normal CRT, and a degraded CRT, and three different noise levels according to the level of noise at full dose, half dose, and quarter dose. The noise of the images had been modified according to the previously measured NPS of a digital mammographic detector and the relationship between dose and noise magnitude.[8] The resolution of the images was modified according to the measured resolution of the medical display devices using a previously verified routine.[6, 9]

3.3   Use the observer model to examine each image and determine the detectability of masses or calcifications in each resolution and noise configuration.
Three different observer models (Non-Prewhitening Matched Filter with Eye Filter (NPWE) observer, the JNDMetrix Visual Discrimination Model, and a Channelized Hotelling Observer (CHO) with Gabor channels) viewed all of the images in this set to determine the detectability of benign masses, malignant masses, and microcalcifications at each noise/resolution configuration. In addition, we examined the impact of resolution and noise on the discriminability between benign and malignant masses. The NPWE and JNDMetrix results were disclosed in Appendix I.

3.4   Develop a fitting method for MTF and NPS curves that reduces the curves to scalar parameters
After obtaining the resolution and noise characteristics, we fit each of them with a multi-parameter exponential function. This provided us with a functional form for the resolution and noise data, which was used by the resolution and noise modification routines.

Task 4: Confirm the empirical model with a small observer performance experiment.

4.1 Create a set of 100 anatomical images from the simulation routine with lesions inserted into 60% of the images.
4.2 Modify the images to the noise and resolution properties of three different digital mammographic systems.
4.3 Perform an Observer Performance experiment with these images with three mammographers
4.4 Compare the mammographers' performance on these images to that predicted by the empirical model. Use statistical power calculation to insure the statistical significance of the results.

As noted in specific aim 2.6 and in Appendix I, the observer model performance differed greatly from human observer performance. Therefore, we merged this specific aim into specific aim 2.3, in which we completed a large scale human observer performance experiment. We increased the number of images viewed in that experiment so that each mammographer rated 2200 images from different resolution and noise levels.

Task 5: Utilize the empirical model to examine the effect of dose on the detection of microcalcifications and masses and determine the minimum allowable dose level for "safe" mammographic imaging.

5.1 Determine the relationship between dose and noise amplitude for the three specific digital mammographic systems through published measurements.
We determined the magnitude of the signal to noise ratio for a given dose by the equation:

$$SNR_{Actual}{}^2 = DQE(0) \cdot SNR_{Ideal}{}^2 \qquad (1)$$

where SNR_{Ideal} was computed using a program by Boone to generate x-ray spectra[10] and $DQE(0)$ was determined from published measurements. This signal to noise ratio was mapped to a graylevel variance using the exposure-pixel value relationship for the detector.

5.2 Determine the effect of scatter utilizing previously published models.
We determined the magnitude of scatter by using previously published data by Boone.[11] Our group measured the magnitude of scatter reduction accomplished by the antiscatter grid. The scatter to primary ratios were then discounted by the scatter reduction from the grid. The effect of scatter was incorporated by reducing the contrast of our simulated lesions by the magnitude of the scattered radiation.

Our previous annual report described our work on scatter using previously published models. Previously published models generally characterize scatter in terms of its magnitude (scatter fraction or scatter to primary ratio). This characterization was appropriate for film-screen systems where scatter primarily affected the contrast of subtle lesions. However, digital systems can overcome these contrast effects, but are still subject to scatter's resolution and noise effects. Therefore, we created a Monte Carlo model of a digital mammographic detector in order to understand scatter's effects. This model is discussed in more detail in Appendix VII.

Using the previously developed empirical model to analyze the effect of dose on the detectability of masses and microcalcifications.
We have generated four image sets using the mammographic data obtained in 1.1. The first set was obtained at full dose and the next three sets have added noise to simulate half, quarter, and eighth dose, respectively. Two different observer models analyzed these image sets, a visual discrimination model and a non-prewhitening matched filter with eye filter model. The results of this work are disclosed in Appendix I. We furthered that work by conducting a large observer experiment with five mammographers. This observer experiment looked at lesion detection and discrimination under three different dose levels. The results of that experiment are disclosed in Appendix II.

5.3   *The results from the previous step will guide the creation for recommendations on the minimum allowable dose for "safe" mammographic imaging.*
As found in Appendix II, observer performance at lesion detection and lesion discrimination remained relatively constant under significant dose reduction. In fact, observer performance did not drop by a statistically significant amount even when the dose was reduced by half. This suggests that the dose in mammography may be reduced modestly with minimal impact on diagnostic performance.[12]

Task 6: Apply the empirical model to ascertain the effect of a specific image processing algorithms, unsharp masking, on lesion detection and optimize its utilization.

6.1   *Examine the clinical parameters used for unsharp masking.*
Several types of unsharp masking are used in clinical practice. We implemented the most basic type of multiscale processing, consisting of an unsharp masking stage and a contrast equalization stage. The form of this processing was determined from previously published methods.[13] The images then underwent a logarithmic transform and were window and leveled for appropriate clinical appearance. The exact parameters for the entire image processing sequence were verified by an experienced breast imaging radiologist, as described in Appendix I and IV.

6.2   *Fit the resolution and noise properties of the combined image processing and detector system using the generalized curve-fitting algorithm.*
The unsharp masking affects the image in the following way:

$$I_S(x,y) = I_O(x,y) + SF \cdot \left(I_O(x,y) - I_O(x,y) \otimes G_B \right) \tag{2}$$

This will affect the image frequency spectra as

$$I_S(u,v) = I_O(u,v) + SF \cdot \left(I_O(u,v) - I_O(u,v) \cdot e^{-(u+v)^2 \sigma^2} \right) ,$$

$$MTF_S(f) = MTF_O(f) \cdot (1 + SF - SF \cdot e^{-f^2 \sigma^2}) \tag{3}$$

where I_S refers to the sharpened image, I_O corresponds to the original image, SF is the sharpness factor, G_B is the Gaussian blurring kernel, *(x,y)* represent the spatial position coordinates, *(u,v)* describe the Cartesian frequency coordinates while *f* refers to the radial frequency coordinate, σ controls the level of blur in the Gaussian kernel, and *MTF* is the modulation transfer function. The MTF

measures the resolution of an image. The noise is best described by the noise power spectra (NPS) and scales as the MTF squared:

$$NPS_S(f) = (1 + SF - SF \cdot e^{-f^2\sigma^2})^2 NPS_O(f)$$ (4)

6.3   *Input the above into the empirical model in order to determine the ideal parameters for unsharp masking which allow for the highest detection levels.*
The results of the human observer experiment, as detailed in Appendix I, suggest that the unsharp masking will not greatly affect observer performance as the image signal to noise ratio will remain constant over this transform.

Task 7: Employ the model to examine the influence of two specific display characteristics, display magnification and display resolution, on lesion detection and thus develop guidelines for optimized viewing of digital mammograms.

7.1   *Determine the effect of display magnification on resolution and noise of an image.*
In the object plane, magnification affects the resolution as:

$$MTF'(u) = MTF(u \cdot m)$$ (5)

where *u* represents spatial frequency in the image plane, *MTF* corresponds to modulation transfer function measured at the detector, *MTF'* is the MTF in the object plane, and *m* refers to the geometric magnification. [14, 15] In the object plane, magnification affects the noise as:

$$NPS'(u) = \frac{1}{m^2} NPS(u \cdot m)$$ (6)

where *NPS'* refers the NPS in the object plane while *NPS* represents the NPS in the image plane.[14, 15]

7.2   *Determine the resolution and noise for four display devices, three common Cathode Ray Tube (CRT) devices and one Liquid Crystal Display (LCD) device.*
Using a high-quality CCD camera, we measured the resolution and noise of two CRT displays and three LCD devices. As LCDs are becoming increasingly common in clinical systems, we decided to include more of a focus on LCD displays than we proposed in the statement of work. The results of this work are in Appendix VI.

7.3   *Fit the resolution and noise properties of the combined display and detector system using the generalized curve-fitting algorithm.*
After obtaining the resolution and noise characteristics, we fit each of them with a multi-parameter exponential function. This provided us with a functional form for the resolution and noise data.

7.4   *Input the above into the empirical model in order to develop guidelines for optimized display of mammographic images.*
Instead of an empirical model, we used a human observer experiment to examine the impact of different display resolutions on the detection of masses and calcifications. Please refer to Appendices I, IV, and IX.

Key Research Accomplishments

- Conducted large observer experiment with five mammographers examining the impact of reduced dose on lesion detection, discrimination, and interpretation time.

- Examined the effect of different medical displays on lesion detection, discrimination, and interpretation time.

- Created large image set with noise properties emulating mammograms acquired at a reduced dose and emulating resolution of images displayed on different commercial medical displays.

- Acquired a large data set of normal digital mammograms.

- Developed Monte Carlo model of digital mammographic system to characterize the effects of x-ray scatter on resolution and noise.

- Developed model for radiographic appearance of breast masses and calcifications and implemented lesion simulation program.

- Measured resolution and noise of five medical displays, representing both CRT and LCD devices.

- Implemented observer model for examining image sets, based on a non-prewhitening matched filter model with eye filter.

- Measured physical characteristics of clinical prototype mammographic system.

- Researched image processing techniques and implemented image processing program.

Reportable Outcomes

Presentations
R.S. Saunders and E. Samei," "Effects of Digital Displays," Southeastern Chapter of the American Association of Physicists in Medicine Educational Symposium, Atlanta, GA (2007). *(Invited Lecture)*

R.S. Saunders, E. Samei, N. Majdi-Nasab, and J.Y. Lo, "Initial Human Subject Results for Breast Bi-Plane Correlation Imaging," SPIE Medical Imaging 2007: Computer Aided Diagnosis (2007).

R.S. Saunders and E. Samei, "Who's afraid of DQE?: How image quality affects clinical performance," Minicourse on Science and Technology for Effective Interpretation of Medical Images, Radiological Society of North America Annual Meeting (2006).

R.S. Saunders and E. Samei, "Displaying Your Health: An Overview of Medical Display Research," American Display Engineering and Applications Conference, Atlanta, GA (2006). (*Invited Lecture)*

R.S. Saunders and E. Samei, "Does Image Quality Impact Mammographic Accuracy?," AAPM 48th Annual Meeting (2006).

R.S. Saunders and E. Samei, "A Monte Carlo Investigation on the Impact of Scattered Radiation on Image Resolution and Noise," SPIE Medical Imaging 2006: Physics of Medical Imaging (2006).

A.S. Chawla, R.S. Saunders, C. Abbey, D.M. Delong, and E. Samei, "Analyzing the effect of dose reduction on the detection of mammographic lesions using mathematical observer models," SPIE Medical Imaging 2006: Image Perception, Observer Performance, and Technology Assessment (2006).

R.S. Saunders, E. Samei, and J. Baker, "Effect of Image Quality on Mammographic Accuracy," Radiology Grand Rounds, Duke University Medical Center, Durham, NC, January 2006. (*Invited Lecture*)

R.S. Saunders, E. Samei, J. Baker, J. Johnson, A. Chawla, and J. Nafziger, "Effect of Image Quality Parameters on the Detection of Mammographic Lesions," Medical Image Perception Society Conference XI (2005).

R.S. Saunders and E. Samei, "Impact of Digital Displays on the Detection of Breast Lesions," Era of Hope Conference (2005).

R.S. Saunders, E. Samei, J. Johnson, and J. Baker, "Effect of display resolution on the detection of mammographic lesions," SPIE Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment (2005).

J.L. Jesneck, R.S. Saunders, E. Samei, J.Q. Xia, and J.Y. Lo, "Detector evaluation of a prototype amorphous selenium-based full field digital mammography system," SPIE Medical Imaging 2005: Physics of Medical Imaging (2005).

R. S. Saunders., E. Samei, and J. Baker, "Simulation of Mammographic Lesions," Radiological Society of North America Annual Meeting (2004).

R. S. Saunders, A. Farshchi, and E. Samei, "Measurement of Display Resolution for Commercial Medical Displays," Radiological Society of North America Annual Meeting (2004).

R. S. Saunders, E. Samei, J. Y. Lo and J. L. Jesneck, "Physical Characterization of a Selenium-based Full Field Digital Mammography Detector," 7th International Workshop on Digital Mammography (2004).

R. S. Saunders, E. Samei, and J. Baker, "Simulation of Breast Lesions," 7th International Workshop on Digital Mammography (2004).

R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," SPIE Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment (2004).

Refereed Journal Articles
R.S. Saunders, J.A. Baker, D.M. Delong, J.P. Johnson, and E. Samei, "Does image quality matter?: Impact of resolution and noise on mammographic task performance," Med. Phys. (Submitted December 2006).

E. Samei, R.S. Saunders, J.A. Baker, D.M. Delong, "Digital Mammography: Effects of Reduced Radiation Dose on Diagnostic Performance," Radiology (In Press, 2007).

R.S. Saunders and E. Samei, "Improving mammographic decision accuracy by incorporating observer ratings with interpretation time," Br. J. Radiol. **79**, 5117-5122 (2006).

R. S. Saunders, E. Samei, J. Baker, D. Delong, M. S. Soo, R. Walsh, E. Pisano, C. M. Kuzmiak and D. Pavic, "Comparison of LCD and CRT Displays Based on Efficacy for Digital Mammography," Acad Radiol **13**, 1317-1326 (2006).

R. S. Saunders, E. Samei, J. A. Baker and D. Delong, "Simulation of Mammographic Lesions," Acad Radiol **13**, 860-870 (2006).

R.S. Saunders, and E. Samei, "Resolution and Noise Measurements of Selected Commercial Medical Displays," Med. Phys. **33**, 308-319 (2005).

R. S. Saunders, E. Samei, J. Y. Lo, J. Jesneck, "Physical Characterization of a Prototype Selenium-based Full Field Digital Mammography Detector," Med. Phys. **32**, 588-599 (2005).

Full-Length Conference Proceeding Articles
R.S. Saunders and E. Samei, "A Monte Carlo Investigation on the Impact of Scattered Radiation on Image Resolution and Noise," Proc. SPIE **6142,** 61423A (2006).

R. S. Saunders, E. Samei, J. Johnson and J. Baker, "Effect of display resolution on the detection of mammographic lesions," Proc. SPIE **5749**, 243-250 (2005).

R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," Proc. SPIE **5372**, 242-250 (2004).

R. S. Saunders Jr., E. Samei, and J. Baker, "Simulation of Breast Lesions," Proceedings of 7th International Workshop on Digital Mammography (2004).

R. S. Saunders Jr., E. Samei, J. Y. Lo and J. L. Jesneck, "Physical Characterization of a Selenium-based Full Field Digital Mammography Detector," Proceedings of 7th International Workshop on Digital Mammography (2004).

Degrees Earned
Robert Saunders graduated with a doctorate in physics from Duke University in May 2006.

Research Opportunities Received based on Training Supported by this Award
Dr. Saunders now serves as a post-doctoral fellow in the Duke Advanced Imaging Labs due to the support of this pre-doctoral grant and the training he received while supported by this pre-doctoral grant.

Funding
Dr. Saunders is currently the recipient of a Komen Postdoctoral Fellowship on the topic: "Biplane Correlation Imaging (BCI) for Early Detection of Breast Cancer." He would not be able to complete this new research without the training he received in breast imaging through this pre-doctoral grant.

List of All Personnel Paid by This Grant
Robert Saunders

Conclusions

This study sought to understand how different imaging parameters affect clinical diagnosis. It proceeded in two stages. First, it developed research tools to measure imaging systems and also tools to simulate breast anatomy and system performance. Using these tools, we measured the physical properties of a clinical mammographic detector and several medical display devices. We also developed a routine that inserts simulated masses or calcifications into a normal mammographic background. Using Monte Carlo methods, we produced a new tool to model scattered radiation in digital radiographic systems. Second, we applied these research tools to several clinically relevant questions. We created a large image set with images emulating those acquired at reduced dose and those displayed at different medical display resolutions. These images were analyzed by computational observer models and by five experienced breast imaging radiologists. The results of these studies addressed two questions. The first question explored the impact of display resolution on the detection of breast masses and calcifications. For this question, we found that different displays had little impact on clinical performance; radiologists performed similarly on all displays. The second question explored the effect of reduced dose on the detection of breast lesions. For this question, we found that the increased noise from reduced dose did impact radiologist performance. However, even reducing the dose by half did not have a statistically significant impact on diagnostic accuracy, suggesting that mammographic dose could be reduced modestly with little impact on clinical performance. Both of these questions have immediate impact on clinical care, as they will determine which medical displays are appropriate for reading mammograms and whether women may be imaged using a lower dose.

References

1. R. Sivaramakrishna, N. A. Obuchowski, W. A. Chilcote and K. A. Powell, "Automatic segmentation of mammographic density," Acad. Radiol. **8**, 250-256 (2001).
2. F. O. Bochud, C. K. Abbey and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," Opt. Express **4**, (1999).
3. R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," Proc. SPIE **5372**, 242-250 (2004).
4. R. S. Saunders Jr., E. Samei and J. A. Baker, "Simulation of Breast Lesions," in Proceedings of 7th International Workshop on Digital Mammography, 162-169 (2004).
5. R. S. Saunders Jr., E. Samei, J. A. Baker and D. Delong, "Simulation of Mammographic Lesions," Acad. Radiol. **13**, 860-870 (2006).
6. R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med. Phys. **30**, 3006-3017 (2003).
7. S. Suryanarayanan, A. Karellas, S. Vedantham, S. M. Waldrop and C. J. D'Orsi, "Detection of simulated lesions on data-compressed digital mammograms," Radiology **236**, 31-36 (2005).
8. S. Suryanarayanan, A. Karellas and S. Vedantham, "Physical characteristics of a full-field digital mammography system," Nucl Instrum Methods **533**, 560-570 (2004).
9. R. S. Saunders and E. Samei, "Resolution and noise measurements of five CRT and LCD medical displays," Med. Phys. **33**, 308-319 (2006).
10. J. M. Boone, T. R. Fewell and R. J. Jennings, "Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography," Med. Phys. **24**, 1863-1874 (1997).
11. J. M. Boone, K. K. Lindfors, V. N. Cooper, 3rd and J. A. Seibert, "Scatter/primary in mammography: comprehensive results," Med. Phys. **27**, 2408-2416 (2000).
12. E. Samei, R. S. Saunders, J. A. Baker and D. Delong, "Digital Mammography: Impact of Reduced Dose on Diagnostic Performance," Radiology (In Press, 2007).
13. M. Stahl, T. Aach and S. Dippel, "Digital radiography enhancement by nonlinear multiscale processing," Med. Phys. **27**, 56-65 (2000).
14. E. Samei, R. S. Saunders, J. Y. Lo, J. T. Dobbins, 3rd, J. L. Jesneck, C. E. Floyd and C. E. Ravin, "Fundamental imaging characteristics of a slot-scan digital chest radiographic system," Med. Phys. **31**, 2687-2698 (2004).
15. S. J. Boyce and E. Samei, "Imaging properties of digital magnification radiography," Med. Phys. **33**, 984-996 (2006).

Appendices

Appendix I
R.S. Saunders, J.A. Baker, D.M. Delong, J.P. Johnson, and E. Samei, "Does image quality matter?: Impact of resolution and noise on mammographic task performance," Med. Phys. (Submitted December 2006).

Appendix II
E. Samei, R.S. Saunders, J.A. Baker, D.M. Delong, "Digital Mammography: Effects of Reduced Radiation Dose on Diagnostic Performance," Radiology (In Press, 2007).

Appendix III
R.S. Saunders and E. Samei, "Improving mammographic decision accuracy by incorporating observer ratings with interpretation time," Br. J. Radiol. **79**, 5117-5122 (2006).

Appendix IV
R. S. Saunders, E. Samei, J. Baker, D. Delong, M. S. Soo, R. Walsh, E. Pisano, C. M. Kuzmiak and D. Pavic, "Comparison of LCD and CRT Displays Based on Efficacy for Digital Mammography," Acad Radiol **13**, 1317-1326 (2006).

Appendix V
R. S. Saunders, E. Samei, J. A. Baker and D. Delong, "Simulation of Mammographic Lesions," Acad Radiol **13**, 860-870 (2006).

Appendix VI
R.S. Saunders, and E. Samei, "Resolution and Noise Measurements of Selected Commercial Medical Displays," Med. Phys. **33**, 308-319 (2005).

Appendix VII
R.S. Saunders and E. Samei, "A Monte Carlo Investigation on the Impact of Scattered Radiation on Image Resolution and Noise," Proc. SPIE **6142**, 61423A (2006).

Appendix VIII
R.S. Saunders Jr., E. Samei, J.L. Jesneck, and J.Y. Lo, "Physical characterization of a prototype selenium-based full field digital mammography detector," Med. Phys. **32**, 588-599 (2005).

Appendix IX
R.S. Saunders Jr., E. Samei, J. Johnson, and J. Baker, "Effect of Display Resolution on the Detection of Mammographic Lesions," Proc. SPIE **5749** (2005).

Appendix X
R. S. Saunders Jr., E. Samei, and J. Baker, "Simulation of Breast Lesions," Proceedings of IWDM (2004).

Appendix XI
R. S. Saunders Jr., E. Samei, J. Y. Lo and J. L. Jesneck, "Physical Characterization of a Selenium-based Full Field Digital Mammography Detector," Proceedings of IWDM (2004).

Does image quality matter?:
Impact of resolution and noise on mammographic task performance

Robert S. Saunders, Jr.
Duke Advanced Imaging Laboratories
Department of Radiology
Duke University Medical Center
Durham, NC 27705

Jay A. Baker
Division of Breast Imaging
Department of Radiology
Duke University Medical Center
Durham, NC 27710

David M. Delong
Department of Biostatistics and Bioinformatics
Duke University
Durham, NC 27710

Jeff P. Johnson
Siemens Corporate Research
755 College Road East
Princeton, NJ 08540

Ehsan Samei
Duke Advanced Imaging Laboratories
Departments of Radiology, Biomedical Engineering, Physics, and Medical Physics
Duke University
Durham, NC 27705

ABSTRACT

The purpose of this study was to examine the effects of different resolution and noise

levels on task performance in digital mammography. This study created an image set

with images at three different resolution levels, corresponding to three digital display

devices, and three different noise levels, with noise magnitudes similar to full clinical

dose, half clinical dose, and quarter clinical dose. The images were read by five

experienced breast imaging radiologists. Human observer results showed increasing

blur had little effect on overall accuracy and individual diagnostic task performance, but

increasing noise caused overall accuracy to decrease by a statistically significant 21%

as the breast dose went to one-quarter of its normal clinical value. The noise effects

were most prominent for the tasks of microcalcification detection and mass

discrimination. The change in accuracy ($\Delta\alpha$) as a function of change in relative

quantum noise magnitude ($\Delta\eta$), or $\Delta\alpha/\Delta\eta$, was -6500 for detection of microcalcifications

and -4200 for discrimination of masses, showing that accuracy at these tasks

decreased strongly as relative quantum noise increased, while $\Delta\alpha/\Delta\eta = 0.88$ for

detection of malignant masses, indicating this task accuracy remained relatively

constant with increasing relative quantum noise. As a secondary aim, the image set

was also analyzed by two observer models to examine whether their performance was

similar to humans. Observer models differed from human observers in their sensitivity

to resolution degradation but were qualitatively similar to human observers in their

sensitivity to noise. The primary conclusions of this study suggest that quantum noise

appears to be the dominant image quality factor in digital mammography, affecting

radiologist performance much more profoundly than display blur.

I.   Introduction

Recent years have seen a rapid expansion in the number of new mammographic imaging technologies, including not only new digital x-ray detectors but also new digital display devices. To rank these new technologies, physicists have conducted a number of studies evaluating the performance of these systems. These studies have generally considered physical parameters such as the resolution, noise, and signal to noise performance of the devices.[1-5] While evaluation of physical parameters provides useful information about a system, this level of evaluation is but one element in considering total system performance.[6-8] Physical metrics do not present a complete picture of system performance because improved physical performance may not lead to improved diagnostic accuracy or, counterintuitively, a degradation of physical performance, such as slightly increased blur or smoothing, may sometimes improve diagnostic accuracy.[9-11] In addition, physical performance results are not always clear. When comparing two systems where one has better resolution while the other has better noise properties, which system's images will be more useful to a radiologist?

As physical metrics present an incomplete picture of system performance, final assessment of a new technology depends on measurements of diagnostic accuracy to learn whether the system helps clinicians discover disease, as this is the main purpose of a screening mammography system.[6-8] Most studies avoid measuring diagnostic performance because the gold standard, human observer studies, consume substantial resources and time compared to physical measurements. One alternative to human observer studies has been using observer models. Observer models are able to

analyze large numbers of images quickly and cheaply compared to human observer experiments. It remains unknown whether these observer models accurately predict human performance and therefore could replace time-consuming human observer performance experiments.

The primary purpose of this work was to establish the connection between physical performance metrics, which are straight-forward to measure, and diagnostic accuracy, which is the purpose of an imaging system. This work specifically examined the diagnostic accuracy of experienced breast imaging radiologists at several mammographic tasks, including the detection and classification of breast lesions. The accuracy of the breast imaging radiologists was measured for several resolution and noise settings in order to learn how accuracy changed as a function of physical image quality. While the primary focus of this study focused on human observer performance, a secondary aim of this paper was to explore whether computational observer models replicated the performance of human observers and could therefore replace human observers in system evaluation.

II.   Methods and Materials

The overall methodology is outlined in figure 1. In summary, a mammographic data set with images at three different resolution levels and three different noise levels was analyzed by human observers. The results were examined to explore the relative impact of different resolution and noise levels on performance. Observer models were

applied to the same image database; their results were compared to the human results to understand how well the observer models track human performance.
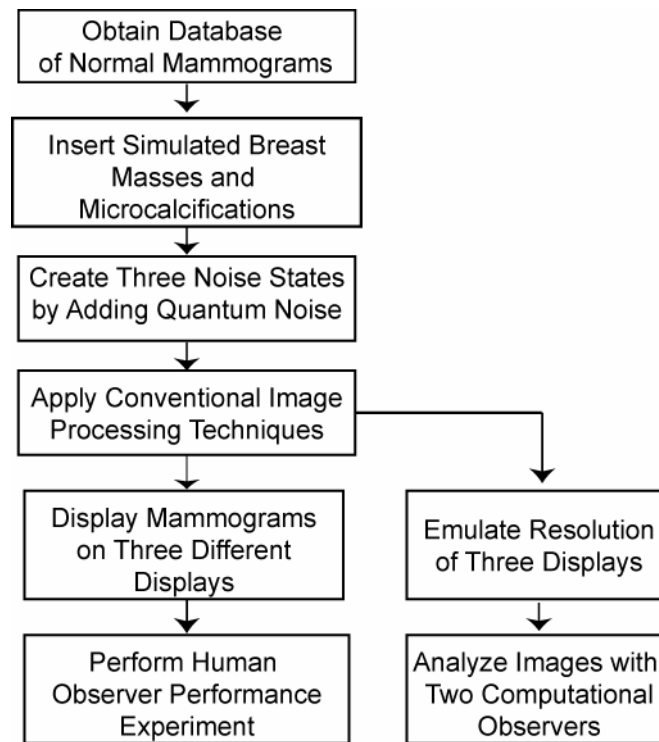

Figure 1. Overview of study methods.

A.   Database of Normal Mammograms

With permission from the Institutional Review Board (IRB), an image database was obtained consisting of 984 normal mammograms acquired on a cesium-iodide based indirect flat-panel detector (GE Senographe 2000D, GE Medical Systems, Waukesha, Wisconsin).[12, 13] From this database, 300 mammograms were selected with the following properties: craniocaudal view, molybdenum anode, and molybdenum or rhodium filtration. The mammograms were acquired with a mean beam energy of 27.6 kVp (σ = 1.42 kVp) and an average compressed breast thickness of 5.13 cm (σ = 1.03

cm). All mammograms were obtained with only gain and dead pixel correction but no additional post-processing by the manufacturer.

B.   Lesion Simulation

To investigate lesion detection and discrimination, simulated mammographic lesions were inserted into the center of normal mammographic regions using a validated lesion simulation routine, the details of which have been previously disclosed.[14-16] This routine relied on measurements from real masses and microcalcifications to guide its simulation. The routine simulated three types of lesions: benign masses, malignant masses, and distributions of microcalcification. The lesion contrast was determined using xSpect software,[17] which modeled the attenuation through a unit thickness of mass or microcalcification accounting for beam energy, tube filtration, breast thickness, and the energy acceptance of the detector. The lesion contrasts were further reduced by the expected scatter fraction for the mammograms.[18] The simulated lesions were scaled by the appropriate contrast and then added to the normal mammogram in a logarithmic manner to maintain the desired contrast independent of the breast attenuation. This process created four categories of images: mammograms with a benign mass, mammograms with a malignant mass, mammograms with a microcalcification distribution, and mammograms with no lesion. There were 150 mammograms in each category. Figure 2 shows example mammographic regions with simulated lesions.
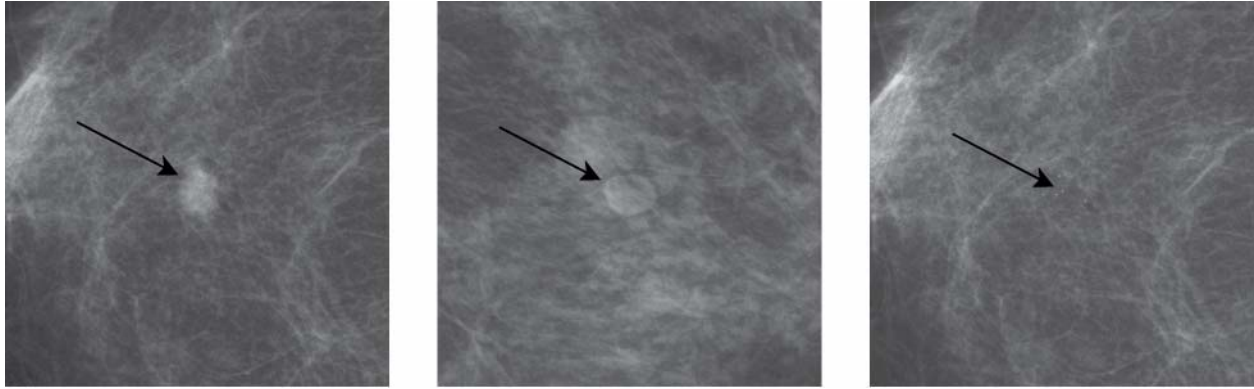
Figure 2. Simulated lesion examples. This includes a malignant mass (left), benign mass (center), and a subtle microcalcification distribution (right).

Pilot studies using observer models showed little change in performance between different resolution and noise levels for the detection of benign and malignant masses but measurable effects for the discrimination of masses and the detection of microcalcifications. Therefore, the mammographic backgrounds for the latter two tasks were paired to increase the statistical power for those tasks. This led to one group of mammograms generating both microcalcification images and normal images (the *microcalcification detection set*) and another class of mammograms generating both benign mass images and malignant mass images (the *mass discrimination set*). Power could have been gained by using the same mammographic backgrounds for all four image categories, but this was not done in order to reduce memory effects in the human observer, which can be a source of bias.

C.   Modification of Quantum Noise in Mammograms

Three levels of quantum noise were employed for this study, as reflected in figure 3, with magnitudes representative of the amount of quantum noise at normal clinical dose ($Noise_1$), half of normal dose ($Noise_2$), and one quarter of normal dose ($Noise_3$). The

normalized noise power spectrum (NNPS) corresponded to the NNPS of the

commercial full-field digital mammography (FFDM) system on which the mammograms
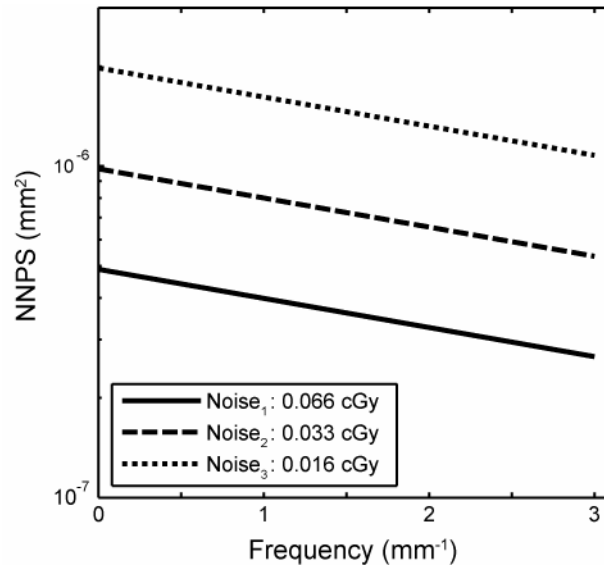
were acquired.[12, 13]

Figure 3(a). Normalized noise power spectrum of the detector noise for an average
mammogram in our database. The relative inherent quantum noise magnitude, or the
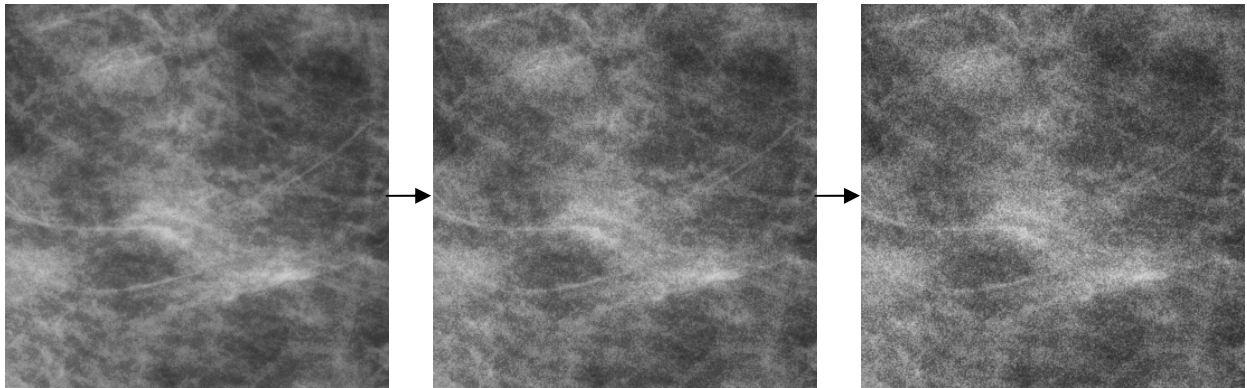integral of the NNPS, was $1.14 \cdot 10^{-5}$, $2.29 \cdot 10^{-5}$, and $4.61 \cdot 10^{-5}$ for $Noise_1$, $Noise_2$, and
$Noise_3$, respectively.

Figure 3(b). Example mammographic regions at each of the noise levels with $Noise_1$
on the left, $Noise_2$ in the center, and $Noise_3$ on the right.

The noise properties of the images were modified by a noise modification routine, the

details of which have been disclosed in a prior publication.[19] The noise modification

routine calculated the signal to noise ratio (SNR) in a mammogram based on the definition of the detective quantum efficiency (DQE) as

$$DQE(0) = \frac{SNR_{Actual}^{2}}{SNR_{Ideal}^{2}}, \tag{1}$$

where *DQE(0)* equals the measured DQE at zero spatial frequency, SNR_{Actual} represents the actual SNR of the mammogram, and SNR_{Ideal} corresponds to the modeled ideal signal to noise ratio. Recognizing that SNR_{Ideal} can be decomposed into *q*, the modeled ideal SNR^2 per unit exposure, multiplied by *ξ*, the measured exposure, equation (1) can be rewritten as

$$SNR_{Actual}^{2} = DQE(0) \cdot SNR_{Ideal}^{2} = DQE(0) \cdot q \cdot \xi. \tag{2}$$

Converting the actual SNR into pixel grayscale units, we obtain the following

$$\frac{\langle x \rangle}{\sigma} = \frac{\sqrt{DQE(0) \cdot q \cdot \xi}}{\mu}, \tag{3}$$

Where *‹x›* corresponds to the mean signal, *σ* refers to the standard deviation of the noise, *‹x›/σ* corresponds to the signal to noise ratio in detector grayscale units and *μ* represents the slope of the measured relationship between exposure and pixel grayscale units. The routine kept the signal level constant and then modified the noise level to produce a mammogram with a signal to noise ratio consistent with a reduced exposure condition.

D.   *Image Processing*

Manufacturers often apply complex image processing algorithms to improve various aspects of image appearance. To control for this process, identical image processing was applied to all mammograms. First, a consistent two-stage image processing

algorithm was applied to accentuate fine detail while equalizing the contrast between the breast and background areas.[20, 21] Second, the grayscale histogram of each mammogram was analyzed to identify the gray level distribution of breast and background. From these results, a window and level was chosen that best balanced the need for contrast at the breast center with visualization of the breast skin line. The window and level function was applied to the mammograms using a sigmoid curve to provide a smooth function over the entire grayscale range. The appropriateness of the algorithm parameters and window and level settings were validated by visual analysis of full mammograms by an experienced radiologist.

E.   Modifying Image Resolution

In this study, the mammograms were evaluated at three resolution levels, illustrated in figure 4. These resolution levels corresponded to the resolution of the entire imaging chain, including the resolution of both the x-ray detector and the digital display. The Modulation Transfer Function (MTF) of the commercial FFDM system on which the mammograms were acquired had been measured by previous investigators.[12, 13] The total resolution combined this detector resolution with the measured resolution of three commercial medical displays: a medical LCD device (*Resolution$_1$*), a medical CRT display with standard resolution (*Resolution$_2$*), and the same CRT but with degraded resolution corresponding to monitor aging (*Resolution$_3$*).[2] The display and detector MTFs were combined assuming the image was displayed with one image pixel per display pixel. The display properties are further described in table I.
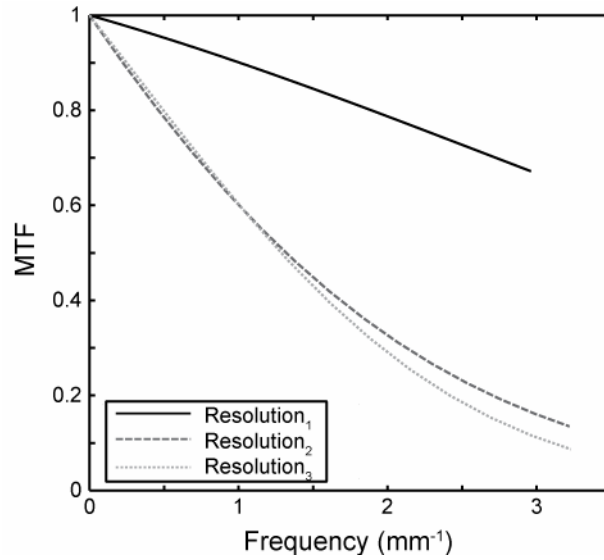
Figure 4(a). Three resolution levels evaluated in this study. These correspond to the resolution of an indirect digital detector convolved with three different display devices. The geometric sharpness, or integral of the MTF^2, was 2.14, 0.92, and 0.90 for *Resolution$_1$*, *Resolution$_2$*, and *Resolution$_3$*, respectively.
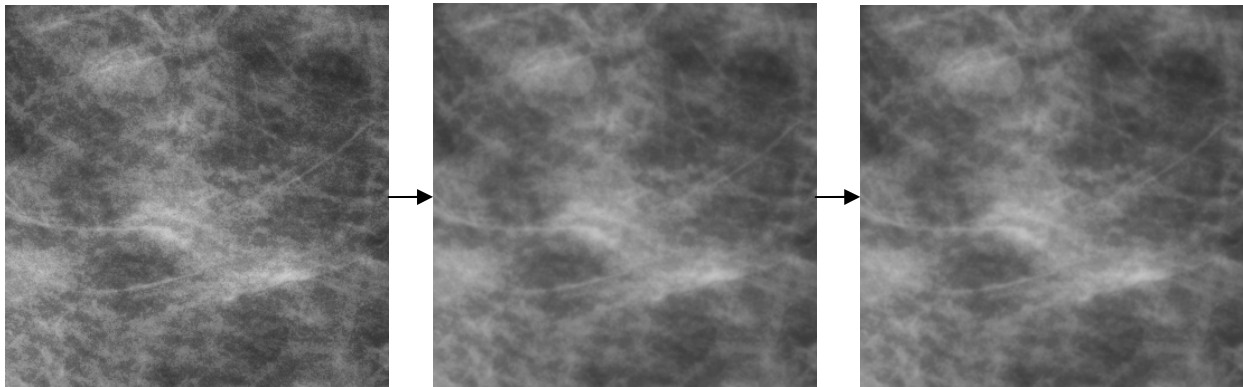

Figure 4(b). Example mammographic regions at each of the three resolution levels with *Resolution$_1$* on the left, *Resolution$_2$* in the center, and *Resolution$_3$* on the right.

Table I: Overview of display properties. The CRT display was further modified by defocusing its electron gun in order to produce both the standard resolution and degraded resolution CRT evaluated in this experiment. The first six rows reflect manufacturer specifications, while the luminance values in the last two rows were measured in our laboratories.[2, 22]

	CRT	LCD
Display Name	Barco MGD 521	National Display Systems Nova V
Display Card	Barco MP1H (10-bit)	RealVision MD5mp (10-bit)
Additional Properties	p45 phosphor	---
Pixel Pitch (mm)	0.148	0.165
Matrix Size	2048 x 2560	2048 x 2560
Active Display Area	304 mm x 380 mm	338 mm x 422 mm
L_{min} (Cd/m^2)	0.52	0.52
L_{max} (Cd/m^2)	308	371

Pilot experiments using observer models guided the allocation of different numbers of images to various resolution and noise subgroups. These pilot experiments showed the smallest effects appeared to occur between different resolution levels. The different resolution categories therefore included additional images to increase the statistical power for finding differences between those categories. Figure 5 illustrates the distribution of images analyzed by observer models and human observers.
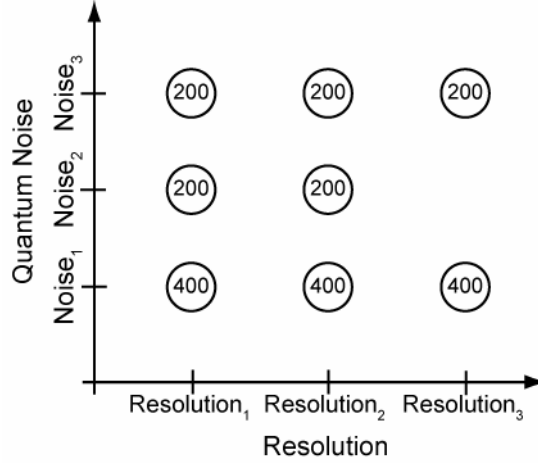
Figure 5. Distribution of images in each resolution and noise category. The category of (Resolution$_3$, Noise$_2$) was not evaluated in this experiment to reduce the duration of the human observer experiment.

The resolution levels were achieved differently for the observer model experiment and human observer experiments. For human observer experiments, the image resolution was altered by displaying the mammograms on three different medical displays with the same desired resolution characteristics. The blur due to the inherent display resolution altered the mammogram resolution to that outlined in figure 4. For the observer model experiments, the image resolution was altered by an established resolution modification routine, the details of which have been disclosed in a prior publication.[19]

F.   Human Observer Performance Experiment

The 2200 images were reviewed by five experienced breast imaging radiologists. The radiologists, from two different academic medical centers, had an average of 11.2 years as a radiology attending, 9.8 years as a mammography attending, and an average reading volume of 160 cases/week. The experiment began with a training set of 100 images, in which feedback was given after each image, in order to familiarize the

radiologist with the lesion types and the graphical user interface (GUI). The radiologists proceeded to the reading set and reviewed 2200 images on a custom GUI which was developed to imitate clinical tasks. The radiologist would view a mammographic region (5.12 cm x 5.12 cm) and rate it into one of four categories: microcalcifications present in the center of the region, a benign mass present in the center of the region, a malignant mass present in the center of the region, or no lesion present. This custom protocol was chosen instead of receiver operating characteristic (ROC) analysis as this protocol increased throughput dramatically, allowing the radiologists to view 2200 images in 2.5—3 hours. The ability to rate this large number of images improved our statistical power and therefore the ability to observe small differences in accuracy between different resolution and noise levels.

To minimize confounding effects, the experiment had several constraints. To maximize image contrast, all images were viewed in a room with low ambient lighting. To minimize an image's rating from being biased by adjacent images, images were shown one at a time with no ability to return to a previously rated image. To accurately reflect display blur, all images were displayed with one display pixel representing one image pixel. To minimize off-axis contrast degradation, the radiologists were asked to view each image straight ahead and centered.[23] To create consistent image appearance, observers could not window and level the images. To minimize various biasing effects, the display order, the image order, and the image orientation were randomized. To minimize fatigue, radiologists were given a five minute break between sessions.

The human observer data were first analyzed for overall classification accuracy and lesion detection accuracy. The overall classification accuracy metric represented the percentage of mammograms correctly rated by an observer. Its associated variance was calculated with a bootstrap analysis, using 10,000 bootstrap samples.[24] Overall lesion detection accuracy was computed as the average of sensitivity and specificity in detecting any lesion. For overall lesion detection accuracy, a true positive was defined as detecting any lesion within an abnormal mammogram, even if the observer misclassifies the lesion as benign or malignant. Its variance was similarly calculated with bootstrap analysis. This detection accuracy metric functioned similarly to the area under an ROC curve as it balanced sensitivity and specificity. Mathematically, the overall lesion detection accuracy can be shown to be the three point approximation to the area under an ROC curve. Statistical significance was estimated using a paired t-test to find a p value.[25]

The data were further analyzed for performance for several clinical tasks: the detection of microcalcifications, the detection of benign masses, the detection of malignant masses, and the discrimination between benign and malignant masses. For each task, the task accuracy examined the ratings for the two related categories. For instance, the two categories analyzed for microcalcification detection were microcalcification images and normal images that were either rated as containing microcalcifications or no lesion. Follow-up experiments confirmed that this exclusion did not bias the task results.[26] Task accuracy was defined as the average of sensitivity and specificity. As before, this accuracy metric was chosen because it approximated the area under an ROC curve.

The task variance was calculated using bootstrap analysis. Task statistical significance was estimated using a paired t-test to find a p value.[25]

G.   Computational Observer Models

1.   Visual Discrimination Model (VDM)

The Sarnoff JNDMetrix[27] Visual Discrimination Model (VDM) was used to predict trends in human performance for four clinical tasks: discrimination of benign and malignant masses and the detection of microcalcifications, benign masses, and malignant masses. This model has been used previously to estimate the detection of mammographic lesions, for example in understanding the effects of display factors and image processing on the detection of mammographic lesions.[28-31] To estimate the detectability of breast lesions, the VDM compared a mammogram containing a lesion to the same mammogram without the lesion and computed a map of just-noticeable difference (JND) values.[32-34] This map was summarized into a scalar value using the Q4 Minkowski normalization technique.[23, 35] We evaluated how this average value changed as a function of resolution and noise properties.

2.   Non-Prewhitening Matched Filter with Eye Filter (NPWE) Model

A non-prewhitening matched filter with eye filter (NPWE) analyzed the images to estimate performance on the same four clinical tasks. The NPWE has been used in several previous investigations to estimate accuracy at mammographic tasks[36-38] and other perception tasks.[39-43] In contrast to a VDM, the NPWE operates on a single image to determine whether it contains a specific lesion.

In summary, the NPWE computed its decision variable by correlating an input image with an observer template. The observer template was formed by filtering the simulated breast lesions by the spatial frequency response of the human visual system *E(f)* using parameters for the eye filter determined from previous publications.[44] The observer templates were scaled to unit standard deviation in order to produce consistent decision variables for each template. The model computed its decision variables in the spatial domain and therefore did not assume stationarity of the backgrounds.[45]

This observer model calculated detection and discrimination using a signal-known-statistically (SKS) paradigm.[46] In this scenario, the location of a lesion within an image is known, but the characteristics of the lesion are only known in a statistical sense. The NPWE model implemented the SKS paradigm by using a bank of lesion templates, which included one template for each simulated lesion template, to represent the lesion characteristics. The NPWE applied the entire filter bank to each image and chose the maximum decision variable. The decision variables were analyzed using Receiver Operating Characteristic (ROC) analysis[47, 48] by computing the non-parametric ROC curve and calculating the area under the non-parametric ROC curve using a trapezoidal numerical integration.

III.   Results

A.   Human Observer Results

Figure 6(a) illustrates the overall classification accuracy of the average human observer at different resolution and noise levels. Overall accuracy differed little between different resolution levels for each noise level. However, overall accuracy dropped substantially as noise increased from $Noise_1$ to $Noise_2$ to $Noise_3$. Figure 6(b) shows lesion detection accuracy at different resolution and noise levels, which shows similar trends to overall classification accuracy. Figure 7 lists human performance at the four clinical tasks. Human observers appeared little affected by resolution for each of four specific clinical tasks. However, human observers did experience accuracy drops for the detection of microcalcifications and the discrimination of masses with increased noise. Increased noise led to a small drop in accuracy for detection of the benign masses at $Noise_3$, but appeared to have a minimal impact on the detection of malignant masses.
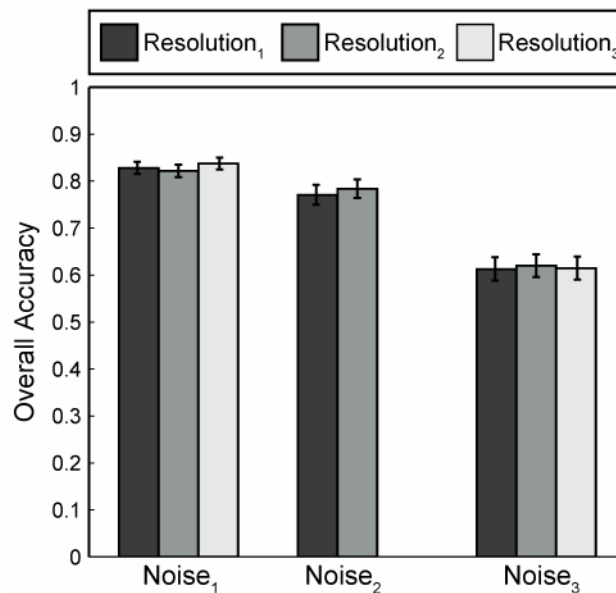
Figure 6(a). Overall classification accuracy for average human observer at different resolution and noise levels.
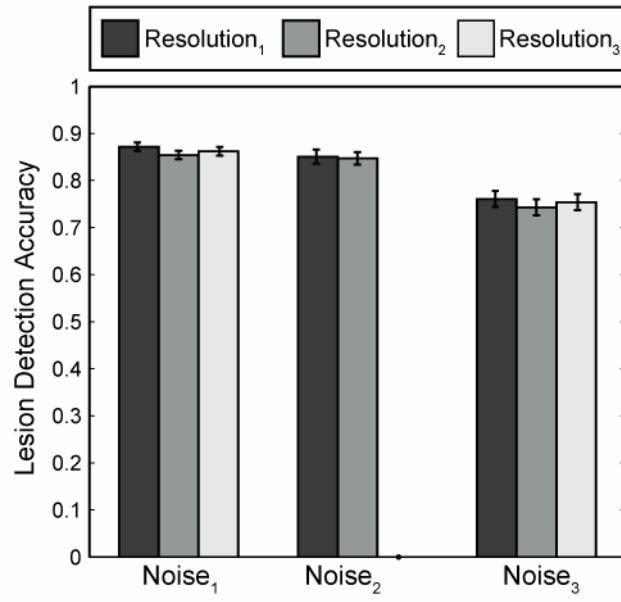
Figure 6(b). Lesion detection accuracy for average human observer at different resolution and noise levels.
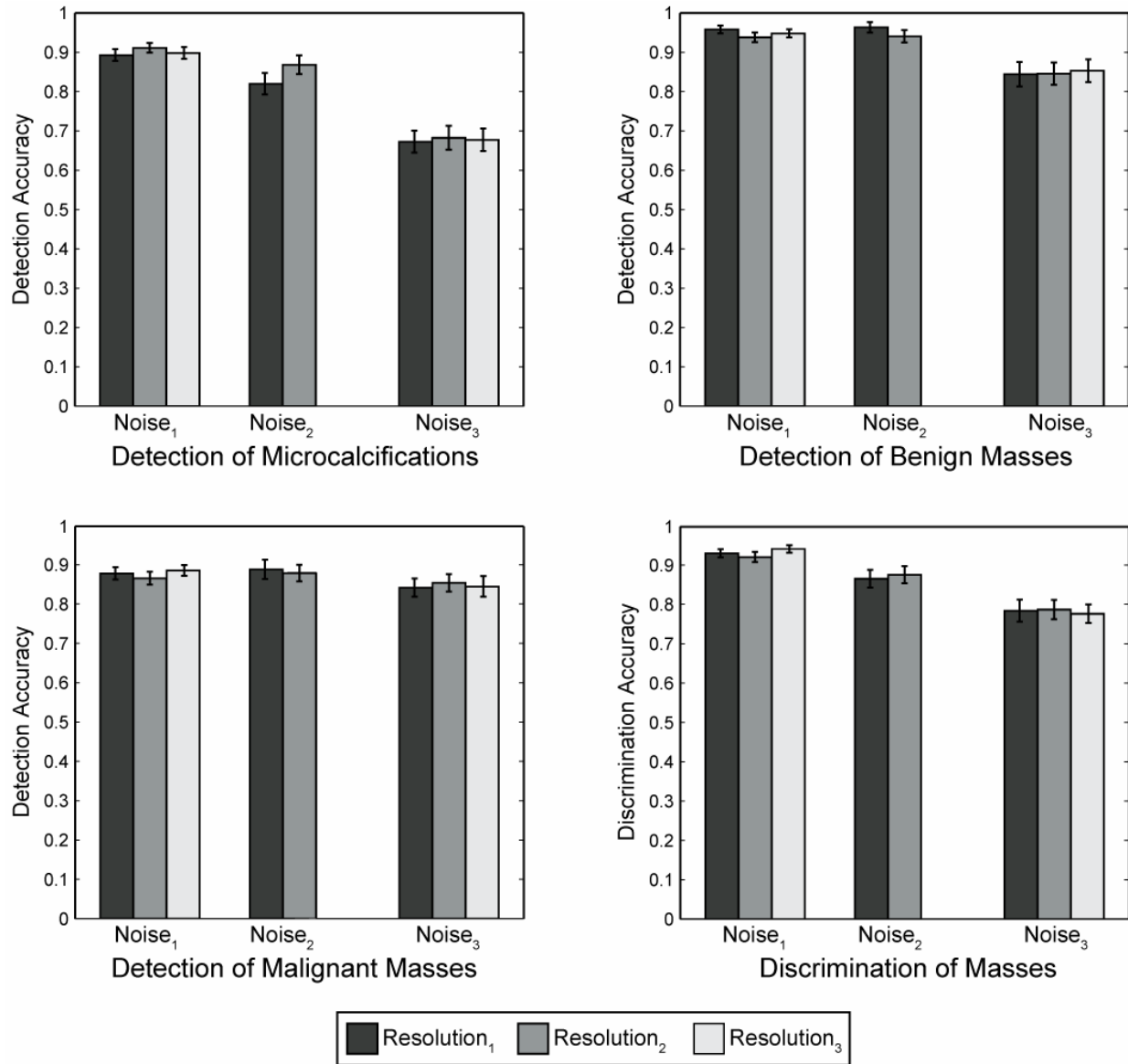
Figure 7. Task performance for average human observer at different resolution and noise levels.

Figure 8 plots overall classification accuracy quantitatively, showing how overall accuracy varies as a function of geometric sharpness, relative quantum noise magnitude, and relative signal to noise ratio. Figure 8(a) quantitatively shows that overall accuracy remained constant with geometric sharpness at each noise level. Figure 8(b) quantitatively shows how overall accuracy decreased with increasing relative quantum noise magnitude. To determine the change in each task accuracy (α) as a function of relative quantum noise magnitude (η), the task accuracies were fit as a function of the relative quantum noise magnitude to find $\Delta\alpha/\Delta\eta$, the slope of the fit line. For the tasks, $\Delta\alpha/\Delta\eta$ was -6500 for detection of microcalcifications, -4200 for discrimination of masses, -3100 for detection of benign masses, and 0.88 for detection of malignant masses. This confirmed that microcalcification detection and mass discrimination had a strong dependence on quantum noise magnitude, while benign mass detection had a weaker dependence and malignant mass detection had minimal relationship to quantum noise magnitude. Figure 8(c) illustrates overall accuracy as a function of inherent signal to noise ratio, confirming that overall accuracy increased as the inherent SNR of the mammogram increased.
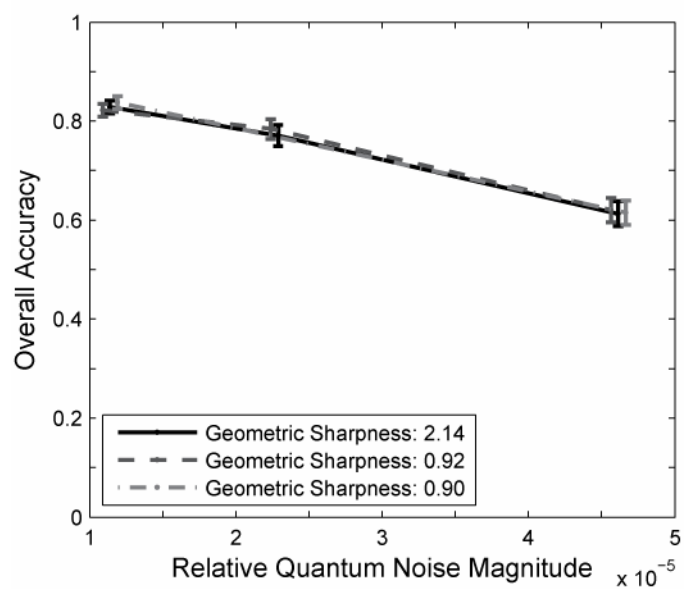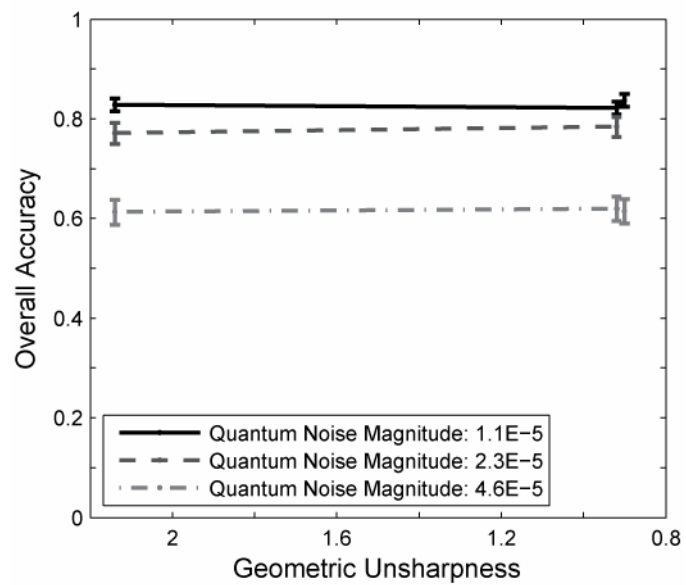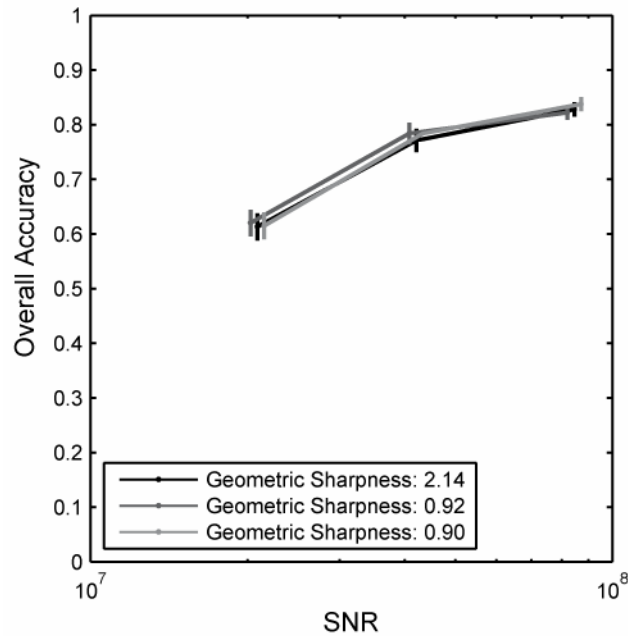
Figure 8. Overall accuracy as a function of geometric sharpness (a), relative inherent quantum noise magnitude (b), and inherent signal to noise ratio (c).

B.   Observer Model Results

1.   Visual Discrimination Model Results

Figure 9(a) illustrates the VDM estimates of mammographic task performance at different resolution levels. Unlike human observers, the VDM shows a drop in performance for all tasks with decreasing geometric sharpness. Figure 9(b) shows the VDM estimates of mammographic task performance at different noise levels. Similar to human observers, the VDM showed little change in performance for detection of masses with increasing noise, but performance for microcalcification detection and mass discrimination decreased with increasing quantum noise. This decrease was not by statistically significant, except for mass discrimination between $Noise_1$ and $Noise_3$ conditions. For all resolution and noise levels, the VDM performed more poorly at microcalcification detection than any other task.
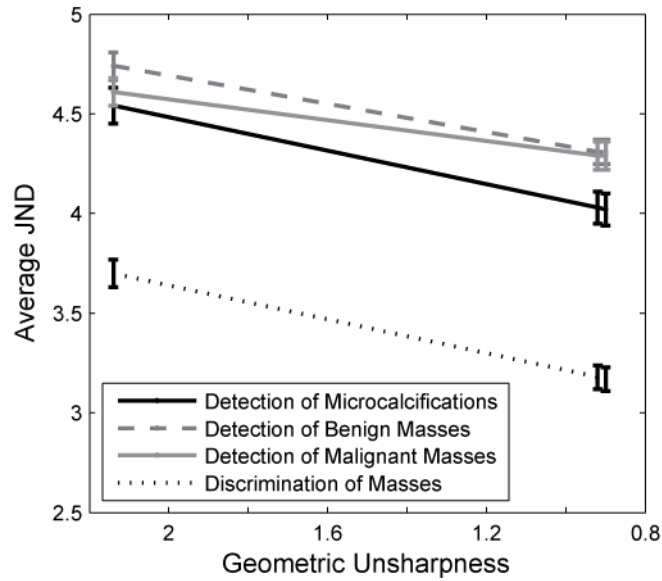
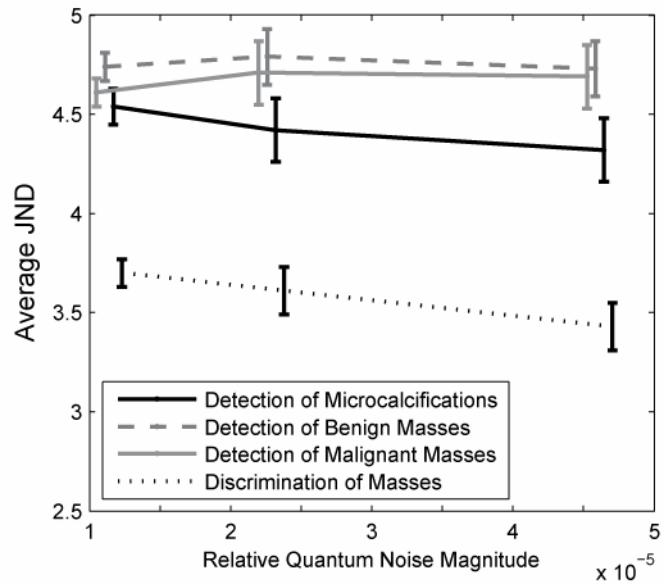**Figure 9(a).** Effect of resolution on task performance using VDM observer.

Figure 9(b). Effect of noise on task performance using VDM observer.

2.   NPWE Results

Figure 10(a) illustrates the NPWE estimates of mammographic task performance at

different resolution levels. Similar to human observers, accuracy remained relatively

constant with decreasing geometric sharpness. Figure 10(b) shows the NPWE estimates of mammographic task performance at different noise levels. Like human observers, performance at mass detection remained similar with increasing noise but, unlike human observers, performance at mass discrimination stayed constant with increasing noise. Analogous to humans, microcalcification detection did decrease slightly with increasing noise. Over all resolution and noise levels, NPWE performance at microcalcification detection was very low, staying near chance levels of 0.5.
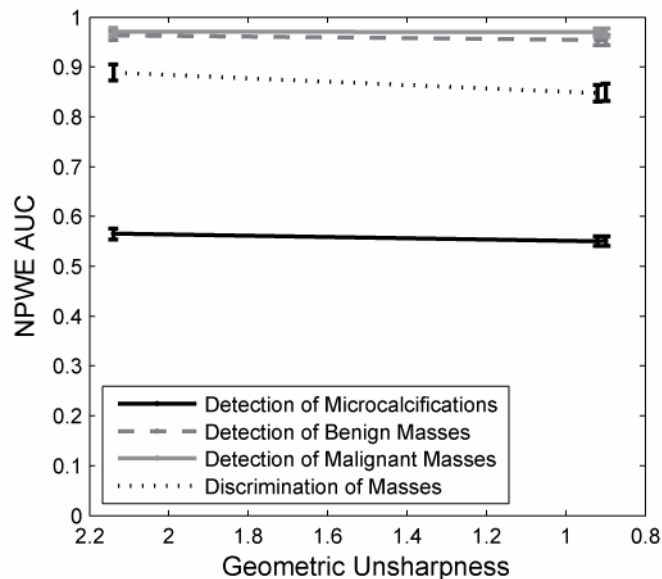

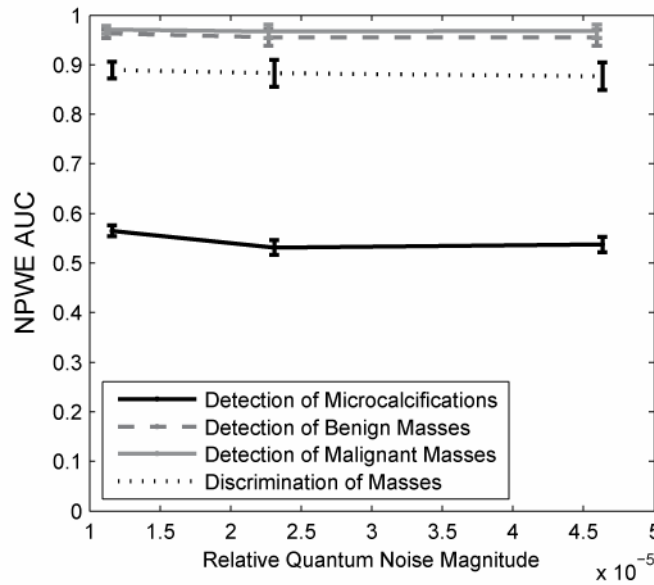**Figure 10(a).** Effect of resolution on task performance using NPWE model

Figure 10(b). Effect of noise on task performance using NPWE model

IV.   Discussion

This work examined mammographic task performance for a number of resolution and noise levels in order to understand how these physical parameters influence the diagnostic utility of images. Human performance was largely unaffected by decreasing resolution but decreased at higher noise levels. Specifically, noise had the greatest effect on human observer performance at microcalcification detection and mass discrimination. In contrast, resolution affected observer model performance strongly while noise had a more modest effect, suggesting that observer models do not yet fully emulate human performance.

Our results showed the importance of noise performance, with resolution playing a more modest role. This confirms earlier work on the detection of simple signals at different

resolutions. Gagne, *et al* used observer models with simple signals and basic backgrounds to study lesion detectability at different resolutions. They found that increased blur could even improve lesion detection within certain ranges.[9, 49] Bacher, *et al* demonstrated similar results with contrast detail experiments with a CDMAM 3.4 phantom. They found similar contrast detail performance for a 5 mega-pixel standard LCD and a CRT.[50] Our work also agrees with previous work on detection at different noise or dose levels. Gagne, *et al* discovered that microcalcification detection decreases at lower doses with roughly a square root dependence on dose.[51] Figure 11 illustrates our microcalcification detection data fit by a square root function, showing that our data roughly follows a square root trend. Roehrig, *et al* examined two early digital mammography systems through a contrast-detail study and found that a system with better noise performance provided a better detection threshold even if it offered lower resolution.[52]
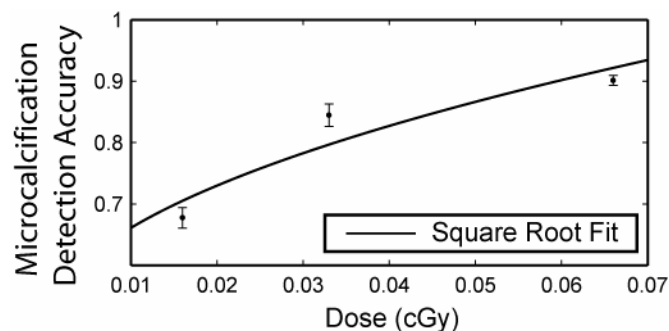

Figure 11. Microcalcification detection accuracy as a function of dose. The solid line illustrates the data fit by a square root function.

Observer model performance did not fully simulate human performance. First, observer model results showed drops in detection and discrimination with increased display blur,

while human observers did not. Computational observers could be more sensitive to display blur than human observers and therefore further work must be conducted to optimize observer models. This difference may also be explained by imperfect simulation of displays, as this study only simulated display resolution and not their noise. However, while the total display noise differs between an LCD and CRT, the amount of perceived display noise should be similar.[2] Second, the NPWE model had difficulty detecting microcalcifications. This deficiency in the NPWE observer has also been noted in a previous study.[53] These points suggest that further work is still needed in optimizing observer models to fully replicate human performance.

While quantum noise and display resolution were plotted as orthogonal variables in figure 5, they are not truly independent. This dependence arises because quantum noise is added before the display blurring step. Display blurring reduces the quantum noise magnitude and its frequency spectra. However, display blurring is not the only way display devices degrade image quality. LCDs possess substantial structured noise due to their pixel structure while CRT devices have luminance variations due to phosphor non-uniformities. This display noise also impacts the total amount of system noise that impedes lesion detection. However, even given these complications, observer performance for each resolution setting remains remarkably similar for each dose setting.

V. Conclusions

This study thoroughly examined the effects of physical measures of image quality on diagnostic accuracy in mammography. One secondary finding of the study was that observer models differed from human observers in their sensitivity to resolution degradation but were qualitatively similar to human observers in their sensitivity to noise. This study found that decreases in resolution by display devices had little impact on human diagnostic performance. However, substantial increases in quantum noise did impede fine-detail tasks, such as the detection of microcalcifications and discrimination of benign and malignant masses. Furthermore, resolution appeared to have little effect at each noise level, suggesting that for this range of resolution and noise parameters, quantum noise may be the dominant image quality factor impeding diagnostic performance.

References

1. R. S. Saunders, Jr., E. Samei, J. L. Jesneck, *et al.*, "Physical characterization of a prototype selenium-based full field digital mammography detector," Med. Phys. **32**, 588-599 (2005).

2. R. S. Saunders and E. Samei, "Resolution and noise measurements of five CRT and LCD medical displays," Med. Phys. **33**, 308-319 (2006).

3. M. B. Williams, P. U. Simoni, L. Smilowitz, *et al.*, "Analysis of the detective quantum efficiency of a developmental detector for digital mammography," Med. Phys. **26**, 2273-2285 (1999).

4. A. D. A. Maidment and M. J. Yaffe, "Analysis of the Spatial-Frequency-Dependent DQE of Optically Coupled Digital Mammography Detectors," Med. Phys. **21**, 721-729 (1994).

5. W. Zhao, W. G. Ji, A. Debrie, *et al.*, "Imaging performance of amorphous selenium based flat-panel detectors for digital mammography: Characterization of a small area prototype detector," Med. Phys. **30**, 254-263 (2003).

6. D. G. Fryback and J. R. Thornbury, "The efficacy of diagnostic imaging," Med. Decis. Making **11**, 88-94 (1991).

7. J. R. Thornbury, "Eugene W. Caldwell Lecture. Clinical efficacy of diagnostic imaging: love it or leave it," AJR. Am. J. Roentgenol. **162**, 1-8 (1994).

8. J. G. Jarvik, "The research framework," AJR. Am. J. Roentgenol. **176**, 873-878 (2001).

9. R. M. Gagne, J. S. Boswell and K. J. Myers, "Signal detectability in digital radiography: spatial domain figures of merit," Med. Phys. **30**, 2180-2193 (2003).

10. A. R. Pineda and H. H. Barrett, "Figures of merit for detectors in digital radiography. I. Flat background and deterministic blurring," Med. Phys. **31**, 348-358 (2004).

11. A. R. Pineda and H. H. Barrett, "Figures of merit for detectors in digital radiography. II. Finite number of secondaries and structured backgrounds," Med. Phys. **31**, 359-367 (2004).

12. S. Suryanarayanan, A. Karellas and S. Vedantham, "Physical characteristics of a full-field digital mammography system," Nucl Instrum Methods **533**, 560-570 (2004).

13. S. Vedantham, A. Karellas, S. Suryanarayanan, *et al.*, "Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype," Med. Phys. **27**, 558-567 (2000).

14. R. S. Saunders Jr., E. Samei and J. A. Baker, "Simulation of Breast Lesions," in Proceedings of 7th International Workshop on Digital Mammography, 162-169 (2004).

15. R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," Proc. SPIE **5372**, 242-250 (2004).

16. R. S. Saunders Jr., E. Samei, J. A. Baker, *et al.*, "Simulation of Mammographic Lesions," Acad. Radiol. **13**, 860-870 (2006).

17. E. Samei and M. J. Flynn, "An experimental comparison of detector performance for direct and indirect digital radiography systems," Med. Phys. **30**, 608-622 (2003).

18. J. M. Boone, K. K. Lindfors, V. N. Cooper, 3rd, *et al.*, "Scatter/primary in mammography: comprehensive results," Med. Phys. **27**, 2408-2416 (2000).

19. R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med. Phys. **30**, 3006-3017 (2003).

20. M. Stahl, T. Aach and S. Dippel, "Digital radiography enhancement by nonlinear multiscale processing," Med. Phys. **27**, 56-65 (2000).

21. A. G. Davies, A. R. Cowen, G. J. S. Parkin, *et al.*, "Optimizing the processing and presentation of PPCR imaging," Proc. SPIE **2712**, 189-195 (1996).

22. E. Samei and S. L. Wright, "Luminance and contrast performance of liquid crystal displays for mammographic applications.," Tech. Canc. Res. Treat. **3**, 429-436 (2004).

23. E. A. Krupinski, J. Johnson, H. Roehrig, *et al.*, "On-axis and off-axis viewing of images on CRT displays and LCDs: observer performance and vision model predictions," Acad. Radiol. **12**, 957-964 (2005).

24. B. Efron and R. Tibshirani, An introduction to the bootstrap (Chapman & Hall, New York, 1993).

25. R. M. Heiberger and B. Holland, Statistical analysis and data display: an intermediate course with examples in S-plus, R, and SAS (Springer, New York, 2004).

26. E. Samei, R. S. Saunders, J. A. Baker, *et al.*, "Digital Mammography: Impact of Reduced Dose on Diagnostic Performance," Radiology (Accepted December 2006).

27. J. Lubin, "A Visual Discrimination Model for Imaging System Design and Evaluation," in *Visual Models for Target Detection and Recognition*, edited by E. Peli (World Scientific Publishers, Singapore, 1995).

28. W. B. Jackson, M. R. Said, D. A. Jared, *et al.*, "Evaluation of human vision models for predicting human observer performance," Proc. SPIE **3036**, 64-73 (1997).

29. E. A. Krupinski, H. Roehrig, M. Engstrom, *et al.*, "MTF correction for optimizing softcopy display of digital mammograms: use of a vision model for predicting observer performance," Proc. SPIE **5034**, 323-327 (2003).

30. E. A. Krupinski, J. Johnson, H. Roehrig, *et al.*, "Use of a human visual system model to predict the effects of display veiling glare on observer performance," Proc. SPIE **5372**, 423-429 (2004).

31. E. A. Krupinski, J. Johnson, H. Roehrig, *et al.*, "Use of a human visual system model to predict observer performance with CRT vs LCD display of images," J. Digit. Imaging **17**, 258-263 (2004).

32. P. J. Jeffrey, L. Jeffrey, S. N. John, *et al.*, "Visual discrimination modeling of lesion detectability," Proc. SPIE **4686**, 248-255 (2002).

33. J. P. Johnson, J. Lubin, E. A. Krupinski, *et al.*, "Visual discrimination model for digital mammography," Proc. SPIE **3663**, 253-263 (1999).

34. R. S. Saunders, E. Samei, J. Johnson, *et al.*, "Effect of display resolution on the detection of mammographic lesions," Proc. SPIE **5749**, 243-250 (2005).

35. S. Winkler, Digital video quality: vision models and metrics (John Wiley & Sons, West Sussex, England, 2005).

36. M. Pachoud, F. O. Bochud, D. Lepori, *et al.*, "Development of a test object for an objective assessment of image quality in conventional or digital mammography," Proc. SPIE **4686**, 219-225 (2002).

37. A. E. Burgess, "Evaluation of detection model performance in power-law noise," Proc. SPIE **4324**, 123-132 (2001).

38. F. O. Bochud, F. R. Verdun, J.-F. Valley, *et al.*, "Importance of anatomical noise in mammography," Proc. SPIE **3036**, 74-80 (1997).

39. A. E. Burgess, "Statistically Defined Background - Performance of a Modified Nonprewhitening Observer Model," J Opt Soc Am A **11**, 1237-1242 (1994).

40. M. P. Eckstein, C. K. Abbey and J. S. Whiting, "Human vs. model observers in anatomic backgrounds," Proc. SPIE **3340**, 16-26 (1998).

41. M. P. Eckstein and C. K. Abbey, "Model observers for signal known statistically tasks," Proc. SPIE **4324**, 91-102 (2001).

42. C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," J Opt Soc Am A **18**, 473-488 (2001).

43. M. P. Eckstein, C. K. Abbey and F. O. Bochud, "Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses," J Opt Soc Am A **17**, 206-217 (2000).

44. M. P. Eckstein, J. L. Bartroff, C. K. Abbey, *et al.*, "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," Opt. Express **11**, 460-475 (2003).

45. F. O. Bochud, C. K. Abbey and M. P. Eckstein, "Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds," J Opt Soc Am A **17**, 193-205 (2000).

46. M. P. Eckstein, Y. Zhang, B. Pham, *et al.*, "Optimization of model observer performance for signal known exactly but variable tasks leads to optimized performance in signal known statistically tasks," Proc. SPIE **5034**, 123-134 (2003).

47. C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**, 720-733 (1986).

48. C. E. Metz, "Fundamental ROC Analysis," in *Handbook of Medical Imaging*, edited by J. Beutel, H. L. Kundel and R. L. Van Metter (SPIE, Bellingham, WA, 2000), Vol. 1, pp 751--770.

49. R. M. Gagne, J. S. Boswell, K. J. Myers, *et al.*, "Lesion detectability in digital radiography," Proc. SPIE **4320**, 316-325 (2001).

50. K. Bacher, P. Smeets, A. De Hauwere, *et al.*, "Image quality performance of liquid crystal display systems: Influence of display resolution, magnification and window settings on contrast-detail detection," Eur. J. Radiol. **58**, 471-479 (2006).

51. R. M. Gagne, B. D. Gallas and K. J. Myers, "Toward objective and quantitative evaluation of imaging systems using images of phantoms," Med. Phys. **33**, 83-95 (2006).

52. H. Roehrig, E. A. Krupinski and T. Yu, "Physical and psychophysical evaluation of digital systems for mammography," Proc. SPIE **2436**, 124-134 (1995).

53.   S. Suryanarayanan, A. Karellas, S. Vedantham*, et al.*, "Detection of simulated lesions on data-compressed digital mammograms," Radiology **236**, 31-36 (2005).

Digital Mammography: Impact of Dose Reduction on Diagnostic Performance

Original Research

Advances in Knowledge:

Decreasing dose in digital mammography by as much as one-half has minimal effect on

the detection of malignant masses but a notable impact on the detection of

microcalcifications, the discrimination between benign and malignant masses, and the

interpretation time.

ABSTRACT

Purpose

To experimentally determine the relationship between radiation dose and accuracy in the detection and discrimination of simulated lesions in digital mammography, using the known simulated lesions as the reference standard.

Materials and Methods

Our HIPAA–compliant study had IRB approval with a waiver of informed consent. Three hundred normal craniocaudal (CC) images were selected from an existing database of digital mammograms. Simulated mammographic lesions mimicking benign and malignant masses and clusters of microcalcifications (3.3-7.4 cm in size) were then superimposed on the images. The images were rendered without and with added radiographic noise simulating the effects of reduced dose by one half and by one quarter of the clinical dose. The images were read by five experienced breast imaging radiologists. The results were analyzed to examine the impact of reduced dose on the overall interpretation accuracy, the detection of microcalcifications, the detection of masses, the discrimination between benign and malignant masses, and the interpretation time.

Results

The overall accuracy dropped from 0.83, to 0.78, to 0.62 for the full, half, and quarter dose levels, respectively. The drop associated with the full-to-quarter transition was statistically significant ($p < 0.01$), primarily due to an effect on the detection of microcalcifications ($p < 0.01$) and the discrimination of masses ($p < 0.05$). That level of

dose reduction did not statistically affect the detection of malignant masses ($p > 0.5$).

However, that increased the mean interpretation time per image by 28% ($p < 0.0001$).

Conclusions

The findings suggest that a reduction of dose in digital mammography has a measurable but modest impact on diagnostic accuracy. The small magnitude of impact in response to the drastic reduction of dose suggests potential for modest dose reductions in digital mammography.

Digital mammography differs from screen-film mammography in important ways (1-4). Foremost, digital mammography captures the image via a digital sensor (5). While in conventional mammography, the analog film serves as both the detector and the display medium, the use of a digital sensor enables a dissociation between the detection and display functions. An important consequence of this dissociation is the independence of display contrast from subject contrast so that that the quality of a digital mammogram is not limited by contrast, which can be manipulated post-acquisition, but rather by noise dictated by the number of photons used to form the image.

The shift from contrast- to noise-limited imaging has a fundamental implication on radiation dose for digital mammography. In clinical implementation of digital mammography, it is imperative to use the appropriate level of radiation (not more and not less) for the diagnostic task at hand. More radiation, on one hand, will lower the level of noise but may impart radiation doses to the patient higher than necessary (6). Less radiation, on the other hand, will lower the signal to noise ratio of the image, which in turn negatively impacts the presentation of the information and thus potentially the diagnosis. The proper level of radiation dose for a mammogram should be dictated by the amount of radiation required to achieve an adequate level of signal-to-noise ratio to present image details required to render an accurate diagnosis.

In the current clinical implementations of digital mammography, the level of radiation dose has generally been set to the dose used by equivalent analog systems. This may partly be due to following the prior convention with analog systems, as well as the fact that the relationship between noise and diagnostic accuracy has not yet been well established for digital mammography. The use of "analog doses" has taken place

despite the fact that the digital systems are not limited by contrast-limited constraints of analog systems, and that the improved detective quantum efficiency (DQE) of most digital systems offers a potential for reduced dose (5, 7). Thus, the purpose of our study was to experimentally determine the relationship between radiation dose and accuracy in the detection and discrimination of simulated lesions in digital mammography, using the known simulated lesions as the reference standard.

Materials and Methods

Our HIPAA-compliant study had IRB approval with a waiver of informed consent.

Image Selection

Three hundred normal craniocaudal (CC) images were randomly selected from an existing database of digital mammograms. All images were originally acquired using a commercial indirect flat-panel mammography system (GE Senographe, GE Medical Solutions, Waukesha, WI) using kVps ranging between 25 and 30 (27.6 kVp, average), a molybdenum anode, and molybdenum or rhodium filtrations. The selected images, considered normal according to the radiologists' reading of the exams in the routine clinical operation, reflected compressed breast thickness ranging from 2.7 to 7.4 cm (5.1 cm, average) and the full range of breast densities from fatty to extremely dense. As a requirement of subsequent steps of the study, the images were utilized in their native raw format without additional post-processing, except for gain and bad pixel corrections implemented by the system.

Simulation of Mammographic Lesions, the Reference Standard

Simulated mammographic lesions, used as the reference standard, were inserted in the selected images using a lesion simulation program (8). Three common categories

of lesions were simulated: benign-appearing masses (modeled after oval circumscribed and oval obscured lesions), malignant-appearing masses (modeled after irregular ill-defined and irregular spiculated lesions), and microcalcifications (modeled after clustered pleomorphic and fine linear branching lesions) (9). All simulated lesions had contrast, contrast profile, shape, border characteristics, and distributions similar to those of the lesion type being simulated. A prior study confirms that radiologists cannot differentiate these simulated lesions from actual lesions (8).

The original set of 300 images was divided into two groups of 150. One group was used to generate both 150 images with benign masses and 150 with malignant masses; the other group was similarly formed into both 150 images with microcalcifications and 150 without lesions. This scheme was designed to enable matching backgrounds for mass discrimination and microcalcification detection tasks to improve the associated statistics, while minimizing the number of times a particular background is viewed by the observers.

The sizes for the simulated lesions were determined based on pilot experiments aiming to target the detection accuracy in the neighborhood of 80% for our experimental condition. The simulated masses ranged in diameter between 3.3 and 4.1 mm (3.7 mm, average). Individual calcifications had mean major and minor axis lengths of 0.37 and 0.25 mm, respectively. The pleomorphic lesions ranged in diameter between 4.0 and 7.0 mm. Distributions for fine linear branching lesions had lines of microcalcifications with lengths between 4.0 and 9.0 mm. The overall contrast magnitudes of the simulated lesions were determined based on the characteristics of image formation and of real lesions (10), and the level of scattered radiation in each mammogram (11). The

simulated lesions were added to the mammographic images in a logarithmic scale to result in contrast magnitudes that would be independent of the image background at the location of the insertion (12).

Noise Addition

As current flat-panel mammography systems are quantum noise-limited (13, 14), the main consequence of a dose reduction is a proportionate increase in the level of quantum noise within the image. Therefore, to create images with a noise appearance similar to that caused by a reduction in radiation dose, a noise modification routine was used to add radiographic noise to the images. To do so, each group of 150 lesion groups described above was divided into three subgroups corresponding to full-dose (without any added noise), half-dose, and quarter-dose of the original (clinical) dose conditions, respectively.

The noise addition routine, previously described in detail (15), was capable of adding noise according to an a priori magnitude and texture (16-19). The desired radiographic noise magnitude was ascertained with the aid of the measured relationship between noise variance and exposure for the imaging system used (Appendix I). At each dose level, the noise magnitude was adjusted based on the pixel value to properly account for the impact of breast attenuation on noise (15). The noise texture was similarly based on the measured noise power spectrum (NPS) for the mammographic system (13-15).

Image Post-processing

The lesion and noise simulation processes described above were performed on the images in the raw format in order to properly emulate the subject contrast of real

mammography lesions and noise properties of low-dose mammograms. However, raw format images were not suitable for interpretation by radiologists. Since GE's proprietary algorithms could not be applied in the post-processing environment, two generic post-processing steps were applied to the images to make the image appearance representative of those in clinical practice. In the first step, unsharp masking and contrast equalization techniques (20) were used to enhance the visualization of smaller structures and to equalize broad signal variations between the center and borders of the breast. The associated parameters for this operation were determined subjectively by visual analysis of processed mammograms (J.A.B., with seven years' experience as a mammography attending, 5,000 cases per year). Identical processing was then applied to all images.

The second processing step established a window and level setting appropriate for optimum viewing of each image. The window and level parameters were determined from histogram analysis of full mammograms, with the goal of clinically-representative contrast levels in the central breast area while maintaining adequate contrast along the breast boundary. A sigmoid function was then fitted to all window and level functions to provide a smooth transition at the extremes of the grayscale range. A breast imaging radiologist (J.A.B., with seven years' experience interpreting mammograms), who did not participate in the subsequent observer study, reviewed all images after window and level processing to ensure the image appearance matched what is common in clinical practice.

Observer Performance Experiment

An observer performance experiment was conducted to assess the impact of reduced dose on lesion detection and discrimination. A 5.12 cm x 5.12 cm (512 pixel x 512 pixel) region, centered at the location of the lesion, was extracted from each image (Figure 1). Using a location-known-exactly experimental paradigm, all images were scored by five breast imaging radiologists with 3-17 (9.8, average) years' experience reading 4000-15000 (8000, average) screening mammograms per year. A custom graphic user interface (GUI) allowed the observers to indicate whether an image appeared to contain a benign mass, a malignant mass, a microcalcification cluster, or no lesion. Observers chose only one answer for each image, and a rating scale was not used. The interface encouraged observers to indicate their choices through the keyboard, substantially shortening the time required for image interpretation. In addition, a modified version of GUI was used for a supplemental experiment in which the observers were only able to score images in terms of a specific diagnostic task (e.g., whether a microcalcification was present or not).

All images were viewed on a 5 mega-pixel liquid crystal display (Nova V, National Display Systems) equipped with a 10-bit display controller (RealVision MD5mp). The device was calibrated to the DICOM grayscale standard display function (21) within 0.52-371 cd/m^2 luminance range (22). All readings were performed in our display laboratory with a controlled low ambient lighting condition.

Before the actual readings, each observer read a different set of 100 images with immediate feedback, to make him/her familiar with the rating interface and appearance of the lesions. Each observer then scored a fixed number of images in each of two sessions, with five-minute breaks between sub-sessions to reduce observer fatigue.

Viewing sessions were 20 to 30 minutes. Two of the observers (eleven and six years'

experience, respectively) did an additional repeat read of the images on the modified

GUI with reduced scoring functionality, which provided scoring options for only the task

at hand, to provide necessary data used to assess the magnitude of potential bias in our

categorical scoring scheme.

Images of different dose levels were displayed in a random order and in one of

six random orientations (4 orthogonal rotations with horizontal/vertical flips) to minimize

reading order and memory effects.

Statistical Analysis

Observer results were analyzed for the effects of varying dose/noise levels on

overall accuracy. Variances were estimated by means of the bootstrap technique

applied over the mammographic images and t-tests were used to compute the statistical

significance of estimated differences, using a Bonferroni correction to preserve type I

error (23, 24). The outcome analyzed was the overall accuracy across all the diagnostic

tasks as represented by two-dimensional contingency tables. Overall accuracy was

computed as the percentage of images correctly rated by each observer and a

combined accuracy statistic was computed as an average over the observers.

While overall accuracy analysis combined all tasks into one figure, the data

analysis further examined the statistical impact of reduced dose on the four specific

clinical tasks, i.e., the detection of microcalcifications, the detection of benign masses,

the detection of malignant masses, and the discrimination between benign and

malignant masses. For each task, a task-specific metric of accuracy was computed as

the average of sensitivity and specificity (Figure 2). This metric is approximately equal to

the area under the bi-normal ROC curve, A_z (25). The task performances were averaged across observers and these results were compared between different dose levels using a t-test for statistical significance with Bonferroni correction (23). To assess the presence of any potential bias associated with our scoring method, the scores from the repeat single-task study were used to adjust the multiple-task data. The standard and adjusted performance on microcalcification detection and mass discrimination were compared to test for any potential bias.

The data analysis included the reading time associated with each dose subgroup of images for each observer, and the average across observers. Standard errors were calculated using bootstrap analyses. As reduced signal to noise ratio might have a detrimental effect on observer confidence which might be reflected in terms of reading time, the data were also analyzed to determine whether the reduced dose images required a longer time for interpretation using survival curves and Proportional Hazard analysis (26, 27).

A p value of less than .05 was considered to indicate a statistically significant difference. All statistical analyses were performed with Matlab Version 7, Release 14 (The Mathworks, Inc., Natick, MA) and JMP 6 (SAS, Cary, NC).

Results

Overall Accuracy

There was a reduction in overall accuracy with reduced radiation dose. While accuracies of individual observers varied, they all exhibited similar trends with reduced dose (Figures 3, 4). The reductions were statistically significant for the full-to-quarter

transition (p < 0.05), and notable but not statistically significant for the full-to-half transition (Table 1).

Task-specific Accuracy

For task-specific average accuracies, there was a clear drop in the detection of microcalcifications and the discrimination between masses with reduced dose at any dose reduction level, with statistical significance associated with the full-to-quarter transition for calcification detection and for mass discrimination. However, the detection of malignant masses did not appear to be much impacted by dose reduction, and the detection of benign masses was only affected when the dose was reduced to a quarter of the normal level (Table 1, Figure 5).

Impact of Bias

We employed a categorical scoring scheme for the observer performance experiment. To assess how this scheme might have biased our task-based results, we also repeated part of the experiment giving observers only binary choices (lesion present vs. lesion absent, or benign vs. malignant). The categorical scoring scheme introduced minimal or no bias, with the results of the two schemes being essentially the same (p > 0.5) (Figure 6).

Timing Performance

Including the training set and breaks, the reading of the entire set of images for each observer took approximately 2.5 hours. The overall timing results, (Figure 7), indicated a discernable impact of dose reduction on the interpretation time. The median interpretation times per image were increased from 2.38 ± 0.07 seconds for full dose, to 2.42 ± 0.09 seconds for half dose to 3.04 ± 0.09 seconds for quarter dose. The

differences were found to be statistically significant (p < 0.0001). The individual observer results confirmed the same behavior, with the average timing performance of the observers appearing to be grossly correlated with their experience and current reading volume.

Discussion

Radiation dose associated with mammographic screening procedures has been a common concern to the radiology community (28-35). In fact, it was in response to such concerns that the US federal government regulated the mammography examinations through the Mammography Quality Standard Act (MQSA) of 1992 (36-38). Recently, there has been an opportunity to potentially reduce the mammographic dose in the transition from analog to digital mammography. Such reductions have been explored by a few studies (31, 39-41). Multiple studies have also indicated the limitations imposed by anatomical noise on mammographic tasks (42-44) which further support such dose reductions. However, concerns about the potential loss of image quality and the resultant impact on diagnostic accuracy have prevented any notable reduction of radiation dose in clinical operations, with clinical implementations still aiming to mostly maintain the dose for digital systems at a level similar to analog systems.

Our study found that decreasing dose in digital mammography by as much as one-half has minimal effect on the detection of malignant masses but a notable impact on the detection of microcalcifications, the discrimination between benign and malignant masses, and the interpretation time. The findings imply that the influence of reduced dose and the associated enhanced noise is mostly in the perception of the high-

frequency components of the lesion signal, as those components represent the defining features of microcalcifications and the distinguishing features that indicate the differences between malignant and benign masses.

The most important clinical implication of our findings confirms that the mammographic dose, even for digital mammography with a potentially higher DQE, has an impact on diagnostic accuracy, and thus proper set up and control of the radiation exposure is an essential requirement for digital mammography operations. However, the small magnitude of the impact in relation to the notable reduction in dose suggests that dose may potentially be decreased with limited impact on clinical utility. That potential is perhaps better appreciated for certain uses such as extra views for images to confirm placement of clips or wires during or after biopsies (45). However, our results imply that there might be a potential for modest reduction of dose in screening applications as well. The confirmation of that implication should await future studies in which accuracy is evaluated at multiple incremental dose levels.

The results of our study are consistent with previous research related to dose reduction in mammography. Dance, et al and Huda, et al used physical measurements to explore the impact of mammographic beam quality and dose reduction on the detection of simple simulated lesions (39, 46). Those studies found that dose could be reduced by using optimum beam qualities while maintaining a constant signal difference to noise ratio. More clinically-based, two additional studies have examined whether reduced dose affects lesion detection by radiologists. Using an indirect flat-panel detector, Obenauer, et al explored the detection of calcifications by imaging an anthropomorphic breast phantom containing simulated calcifications (41). Similarly,

Hemdal, et al conducted a human performance experiment using 28 real mammograms acquired at full and half dose (31). Both studies found potential for substantial dose reduction for digital mammography. These prior studies have either been limited by the physical measures of image quality (i.e., signal difference to noise ratio) or the limited number of clinical tasks and lesion types examined. Our study, in contrast, explored a much larger number of clinical tasks, employing a greater number of images and lesions, allowing the results to be more generalizable to a larger patient population.

Most diagnostic observer performance experiments are currently based on rating of images for the presence of a single type of abnormality into multiple grades, ranging from definitely absent to definitely present with multiple grades in between. The number of gradations range between 4 to 100 (47-49). While this approach is essential for ROC analysis (50, 51), the current de facto standard for evaluating diagnostic systems, it falls short of reflecting many diagnostic tasks performed in the clinic today, when an examiner needs to make binary decisions about the presence or need for a biopsy for a multiplicity of abnormalities that might be depicted by an image. In our study, we asked the observers to rate images for the presence of different types of abnormalities without confidence ratings. This categorical approach closely emulated the clinical paradigm. It also substantially shortened the time required for rating an individual image.

While the above approach has a strong appeal in terms of clinical relevance, it might be prone to potential biasing problems. A bias might be introduced when the assessment of a given diagnosis is impacted by the inclusion of a rating that is not relevant to the task at hand. For example, when assessing the discrimination of benign and malignant masses on an image, an observer might change his/her natural score if

an option is provided for scoring for the presence of microcalcifications (i.e., an unrelated option). If that change is more or less for malignant mass images than for benign mass images, that would create a bias in the results. Recognizing that this might have a potential impact on our results, we performed a supplemental study in which two observers were provided with only scoring options relevant to the task at hand. A comparison of the results with and without the multiple-scoring option indicated that a potential impact of bias was non-existent, at best, or minimal, at worst. The findings encourage the use of categorical methodologies for future observer performance experiments.

Our study has limitations. First, while the results indicate the relative impact of dose reduction on various diagnostic tasks in digital mammography, the direct relationship of breast dose and diagnosis could only be inferred as the reduction was applied only in a relative sense: For a given image acquired at a specific radiographic technique, breast dose is directly related to exposure and noise and thus a relative reduction in dose can be achieved by a linear reduction of exposure and a corresponding increase in radiographic noise. However, the relationship between exposure, dose, and noise is dependent on the kVp, beam filtration, breast composition, and breast thickness, which vary from image to image. Thus, while our results can tell us what would happen if for a given breast a lower than standard mAs or exposure is used, they do not tell us the specific quantitative relationship between glandular breast dose and accuracy. Secondly, our study investigated the impact of dose reduction using a signal-known exactly paradigm in which the observers knew the approximate location of a lesion. This strategy, while eliminating visual search, was implemented to keep

other sources of variability under control. However, if we had used the full images and incorporated search, we might have possibly observed bigger differences as a function of dose, a prospect which cannot be substantiated with our results. Finally, our study was based on simulated lesions and dose levels. While the simulations were realistic, there are always differences between real and simulated situations, which might have a bearing on the findings.

In summary, the findings of our experimental study suggest that a reduction of radiation dose by as much as one-half can have a measurable but modest impact on diagnostic accuracy in digital mammography, particularly in the detection of microcalcifications and the discrimination between malignant and benign masses. The dose reduction also appears to lengthen the interpretation time.

Practical Application

The results suggest that, given the small magnitude of impact on accuracy in response to the drastic reduction of dose, there may be a potential for modest dose reductions in digital mammography. While that potential awaits a confirmation by a follow-up clinical trial, careful attention should be paid to utilized radiation dose and associated image quality when setting up and operating digital mammography units.

Appendix

To accurately simulate dose reduction in a mammogram, the magnitude of the added noise needs to correspond with a proportionate reduction of exposure. In our study, we maintained the mean pixel value of the images, but altered the image signal-to-noise ratio (SNR) to simulate the effects of reduced exposure. The actual SNR of an image is related to the Detective Quantum Efficiency (DQE) and the ideal SNR ratio as

$$DQE(f=0) = \frac{SNR_{Actual}^{2}}{SNR_{Ideal}^{2}} = \frac{SNR_{Actual}^{2}}{q \cdot \xi},$$

where q represents the ideal SNR squared per unit exposure and ξ is the exposure (52). Using measured values of the DQE for the mammographic detector (13, 14) and the estimated values for the ideal SNR, calculated using an x-ray modeling program (xSpect, Henry Ford Health System, Detroit, Michigan) (10), this equation was solved to determine the actual SNR at different exposure levels. The scalar magnitude of the noise was then determined from the computed SNR values using

$$\sigma_{AdditionalNoise} = \xi_{In} \sqrt{SNR_{Actual}\big|_{FullDose} - SNR_{Actual}\big|_{ReducedDose}}$$

where σ indicates the standard deviation of the added noise, and ξ_{in} is the exposure associated with the input image being modified.

Acknowledgements

References

1. Fischer U, Hermann KP, Baum F. Digital mammography: current state and future aspects. Eur Radiol 2006; 16:38-44.

2. Conant EF, Maidment AD. Update on digital mammography. Breast Dis 2001; 13:109-124.

3. James JJ. The current status of digital mammography. Clin Radiol 2004; 59:1-10.

4. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353:1773-1783.

5. Yaffe MJ, Mainprize JG. Detectors for digital mammography. Technol Cancer Res Treat 2004; 3:309-324.

6. BEIR. Health risks from exposures to low levels of ionizing radiation. In. Wahington, DC: National Academies of Science, 2005.

7. Haus AG, Yaffe MJ. Screen-film and digital mammography. Image quality and radiation dose considerations. Radiol Clin North Am 2000; 38:871-898.

8. Saunders R, Samei E, Baker J. Simulation of mammographic lesions. Acad Radiol 2006; 13:860-870.

9. C. D'Orsi ea. Illustrated Breast Imaging Reporting and Data System (BI-RADS). Reston, VA: American College of Radiology, 1998.

10. Samei E, Flynn MJ. An experimental comparison of detector performance for direct and indirect digital radiography systems. Med Phys 2003; 30:608-622.

11. Boone JM, Lindfors KK, Cooper VN, 3rd, Seibert JA. Scatter/primary in mammography: comprehensive results. Med Phys 2000; 27:2408-2416.

12. Samei E, Flynn MJ, Eyler WR. Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. Radiology 1999; 213:727-734.

13. Suryanarayanan S, Karellas A, Vedantham S. Physical characteristics of a full-field digital mammography system. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 2004; 533:560-570.

14. Vedantham S, Karellas A, Suryanarayanan S, et al. Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype. Med Phys 2000; 27:558-567.

15. Saunders RS, Samei E. A method for modifying the image quality parameters of digital radiographic images. Med Phys 2003; 30:3006-3017.

16. Giger ML, Doi K, Fujita H. Investigation of Basic Imaging Properties in Digital Radiography .7. Noise Wiener Spectra of II-TV Digital Imaging-Systems. Med Phys 1986; 13:131-138.

17. Giger ML, Doi K, Fujita H. Analysis of Noise Wiener Spectra in Digital Ii/Tv Imaging-Systems. Med Phys 1984; 11:385-385.

18. Giger ML, Doi K, Metz CE. Investigation of Basic Imaging Properties in Digital Radiography .2. Noise Wiener Spectrum. Med Phys 1984; 11:797-805.

19. Saunders RS, Jr., Samei E, Jesneck JL, Lo JY. Physical characterization of a prototype selenium-based full field digital mammography detector. Med Phys 2005; 32:588-599.

20. Stahl M, Aach T, Dippel S. Digital radiography enhancement by nonlinear multiscale processing. Med Phys 2000; 27:56-65.

21. NEMA-PS3.14. Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Display Standard Function. In. Rosslyn, VA: National Electrical Manufacturers Association, 2000.

22.   Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report. Med Phys 2005; 32:1205-1225.

23.   Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol 2001; 54:343.

24.   Casella G, Berger RL. Statistical inference. Australia; Pacific Grove, CA: Thomson Learning, 2002.

25.   Swets JA, Pickett RM. Evaluation of diagnostic systems : methods from signal detection theory. New York: Academic Press, 1982.

26.   Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 1972; 34:187-220.

27.   Lawless JF. Statistical models and methods for lifetime data. New York: Wiley, 1982.

28.   Parker MS, Hui FK, Camacho MA, Chung JK, Broga DW, Sethi NN. Female Breast Radiation Exposure During CT Pulmonary Angiography. AJR Am J Roentgenol 2005; 185:1228-1233.

29.   Morin Doody M, Lonstein JE, Stovall M, Hacker DG, Luckyanov N, Land CE. Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. Spine 2000; 25:2052-2063.

30.   Sigurdson AJ, Doody MM, Rao RS, et al. Cancer incidence in the US radiologic technologists health study, 1983-1998. Cancer 2003; 97:3080-3089.

31.   Hemdal B, Andersson I, Grahn A, et al. Can the average glandular dose in routine digital mammography screening be reduced? A pilot study using revised image quality criteria. Radiat Prot Dosimetry 2005; 114:383-388.

32. Ramos M, Ferrer S, Villaescusa JI, Verdu G, Salas MD, Cuevas MD. Use of risk projection models to estimate mortality and incidence from radiation-induced breast cancer in screening programs. Phys Med Biol 2005; 50:505-520.

33. Law J, Faulkner K. Two-view screening and extending the age range: the balance of benefit and risk. Br J Radiol 2002; 75:889-894.

34. Law J, Faulkner K. Concerning the relationship between benefit and radiation risk, and cancers detected and induced, in a breast screening programme. Br J Radiol 2002; 75:678-684.

35. Huda W, Sourkes AM, Bews JA, Kowaluk R. Radiation doses due to breast imaging in Manitoba: 1978-1988. Radiology 1990; 177:813-816.

36. Suleiman OH, Spelic DC, McCrohan JL, Symonds GR, Houn F. Mammography in the 1990s: the United States and Canada. Radiology 1999; 210:345-351.

37. Pisano ED, Schell M, Rollins J, et al. Has the mammography quality standards act affected the mammography quality in North Carolina? AJR Am J Roentgenol 2000; 174:1089-1091.

38. Kneece J. Breast imaging: why MQSA (Mammography Quality Standards Act). Adm Radiol 1994; 13:33-34.

39. Huda W, Sajewicz AM, Ogden KM, Dance DR. Experimental investigation of the dose and image quality characteristics of a digital mammography imaging system. Med Phys 2003; 30:442-448.

40. Dance DR, Thilander AK, Sandborg M, Skinner CL, Castellano IA, Carlsson GA. Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: a Monte Carlo study. Br J Radiol 2000; 73:1056-1067.

41. Obenauer S, Hermann KP, Grabbe E. Dose reduction in full-field digital mammography: an anthropomorphic breast phantom study. Br J Radiol 2003; 76:478-482.

42. Bochud FO, Valley JF, Verdun FR, Hessler C, Schnyder P. Estimation of the noisy component of anatomical backgrounds. Med Phys 1999; 26:1365-1370.

43. Bochud FO, Verdun FR, Valley J-F, Hessler C, Moeckli R. Importance of anatomical noise in mammography. Proc SPIE 1997; 3036:74.

44. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. Med Phys 2001; 28:419-437.

45. Riedl CC, Jaromi S, Floery D, Pfarl G, Fuchsjaeger MH, Helbich TH. Potential of dose reduction after marker placement with full-field digital mammography. Invest Radiol 2005; 40:343-348.

46. Analoui M, Krisnamurthy S, Brown C. Modeling and measurement of root canal using stereo digital radiography. Proc SPIE 2000; 3976:306-314.

47. Metz CE. Fundamental ROC Analysis. In: Beutel J, Kundel HL, Van Metter RL, eds. Handbook of Medical Imaging. 1 ed. Bellingham, WA: SPIE, 2000; 751--770.

48. Chesters MS. Human visual perception and ROC methodology in medical imaging. Phys Med Biol 1992; 37:1433-1476.

49. Turner SR, Samei E, Hertzberg BS, et al. Sonography of fetal choroid plexus cysts: detection depends on cyst size and gestational age. J Ultrasound Med 2003; 22:1219-1227.

50. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989; 24:234-245.

51. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986; 21:720-733.

52. Dobbins J. Image Quality Metrics for Digital Systems. In: Beutel HKJ, Metter RV, eds. Handbook of Medical Imaging. Washington, DC: SPIE, 2000; 163-222.

Tables

Table I. Summary statistics of the results indicating the reduction in accuracy in average observer performance. Positive values correspond to a reduction and negative values to an enhancement. Statistically significant transitions at 95% confidence level ($p < 0.05$) are indicated with a star.

Task	Full- to half-dose transition	Full- to quarter-dose transition
Overall Accuracy (all tasks combined)	0.05	0.21[*]
Accuracy in the detection of micro-calcifications	0.06	0.22[*]
Accuracy in the detection of benign masses	0.00	0.10
Accuracy in the detection of malignant masses	-0.01	0.02
Accuracy in the discrimination between malignant and benign masses	0.05	0.14[*]

Figure Captions

Fig. 1: Examples of mammographic images used for the observer experiment at full dose (first column), half dose (second column), and quarter dose (third column) with microcalcification distributions (first row), malignant masses (second row), and benign masses (third row).

Fig. 2: An example contingency table illustrating its use to deduce performance results for the example task of the detection of malignant masses.

N = normal

C = microcalcification

B = benign mass

M = malignant mass

Fig. 3: The contingency tables at the three dose levels, full dose (a), half dose (b), and quarter dose (c), averaged across observers, indicating the fraction of which the observers scored the images of a given class.

Fig. 4: Variation in the overall accuracy, representing the average of all the diagnostic tasks involved, as a function of dose level for individual observers and the average across observers. The variance for each observer was calculated using bootstrap analysis, with error bars representing one standard deviation. The figure illustrates that

for each observer and across all observers, overall accuracy is reduced as radiation dose decreases.

Fig. 5: The impact of dose level on the detection of microcalcifications, malignant masses, and benign masses, and the discrimination of malignant and benign masses. The bar data correspond to the averages from all observers, with error bars calculated in a similar fashion as in Figure 4. With the full-to-quarter dose reduction, there was a significant decrease in calcification detection and mass discrimination. The detection of malignant masses was reduced only at the one-quarter dose level, and the detection of benign masses changed little when radiation dose was reduced.

Fig. 6: The potential impact of bias in the detection of microcalcifications associated with multiplicity of observer grading tasks illustrated with the results acquired with potential bias and adjusted to remove such potential bias. Error bars represent one standard deviation. Nearly identical results were found for the two scoring schemes, categorical and two-task, illustrating that the categorical scoring introduced minimal or no bias.

Fig. 7: Number of images unrated by a given time. The three lines compare the reading times for images with signal to noise ratios reflective of full clinical dose, half dose, and quarter dose. A statistically significant relationship was found between radiation dose and observer interpretation time.
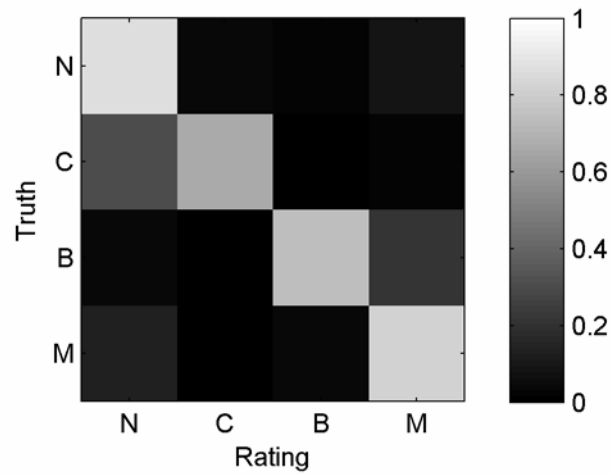
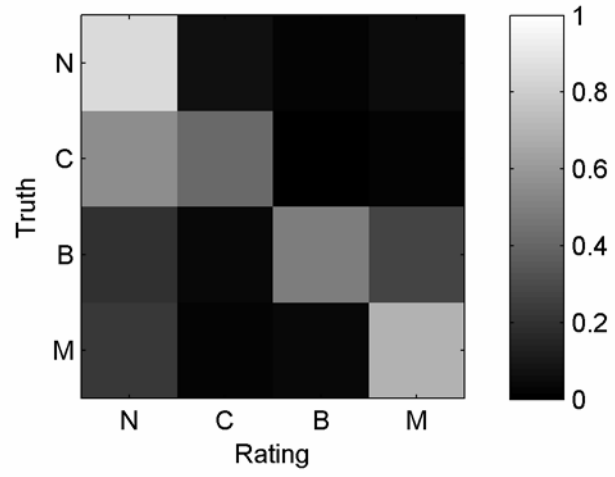**Figures**
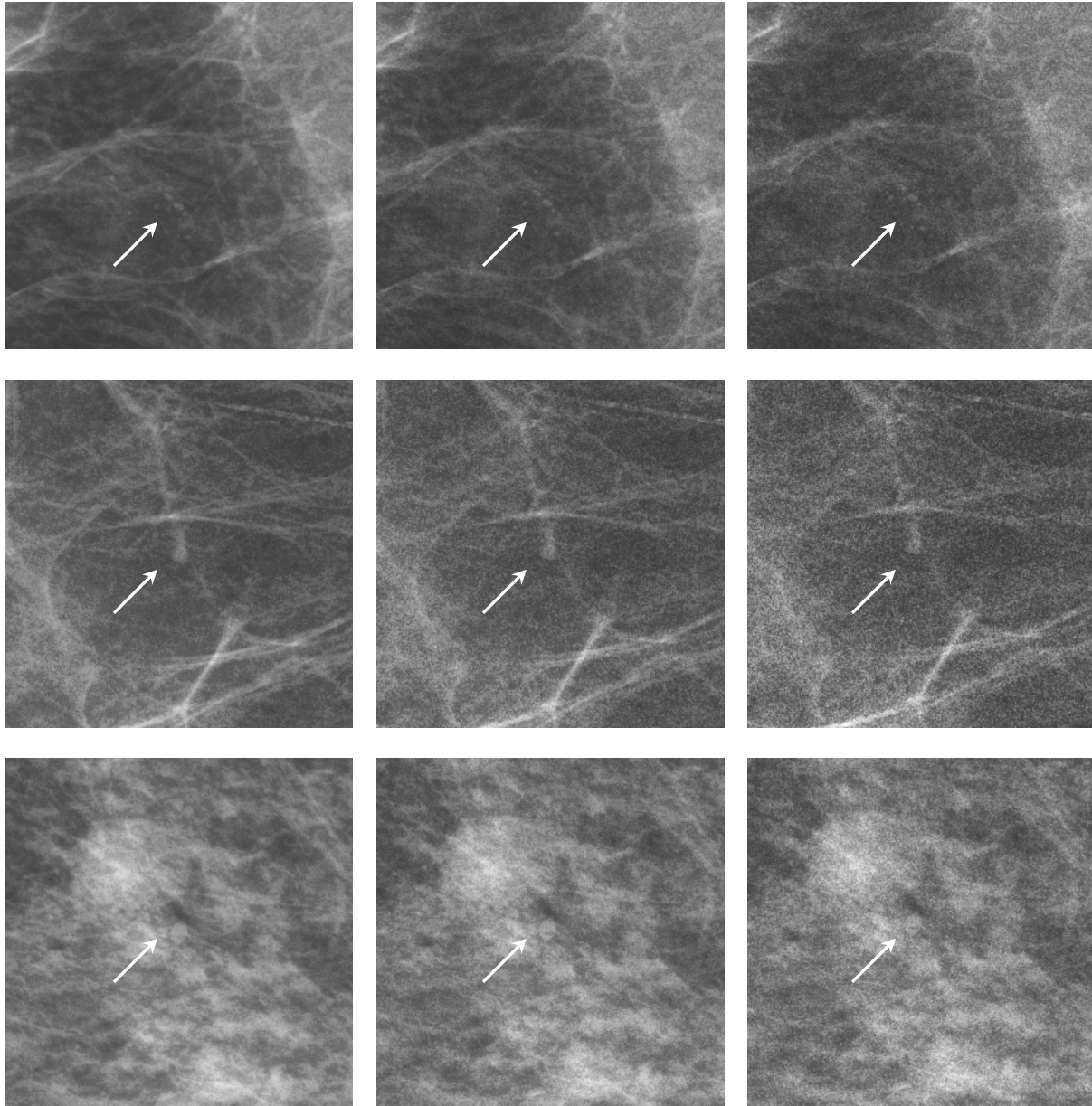
Figure 1
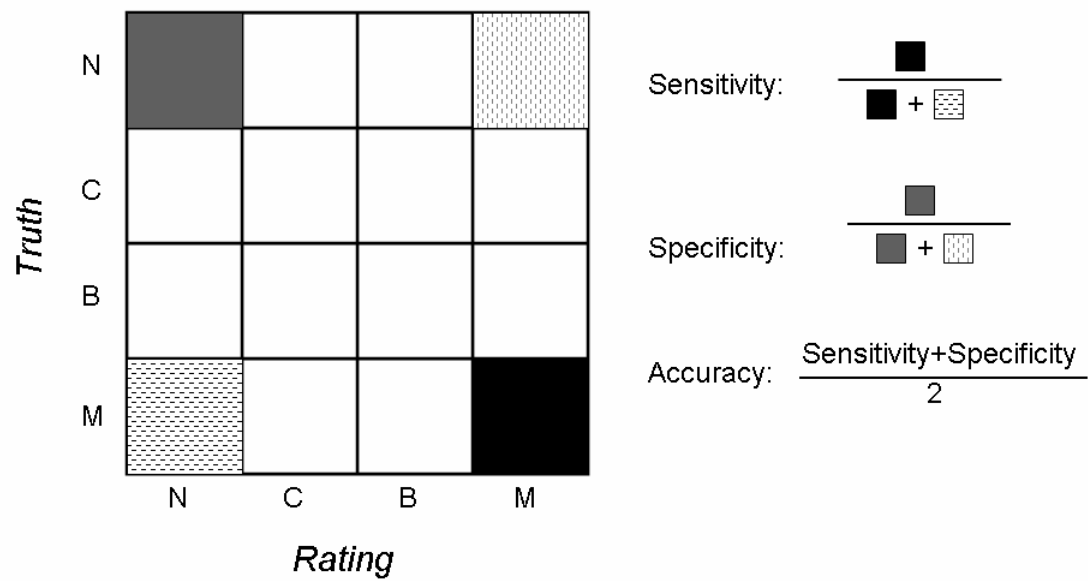
Figure 2
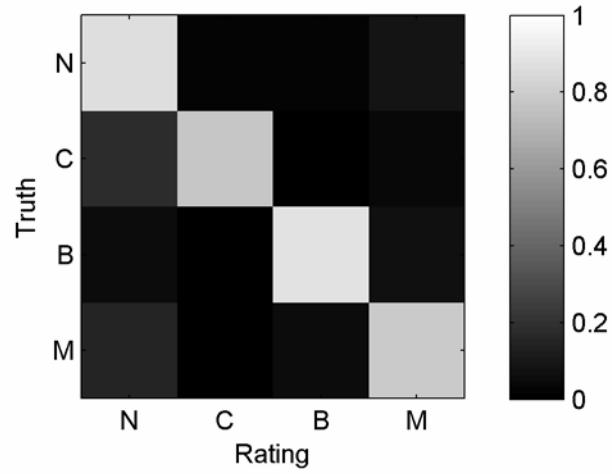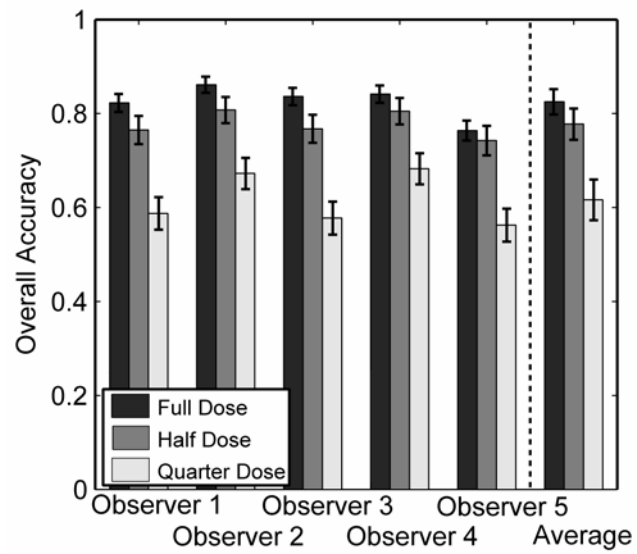
(a) Full Dose Contingency Table
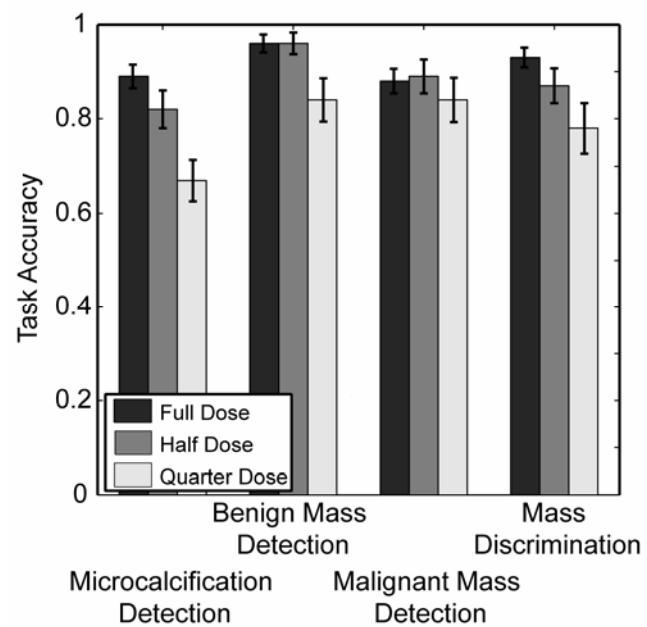

(b) Half Dose Contingency Table

(c) Quarter Dose Contingency Table

Figure 3

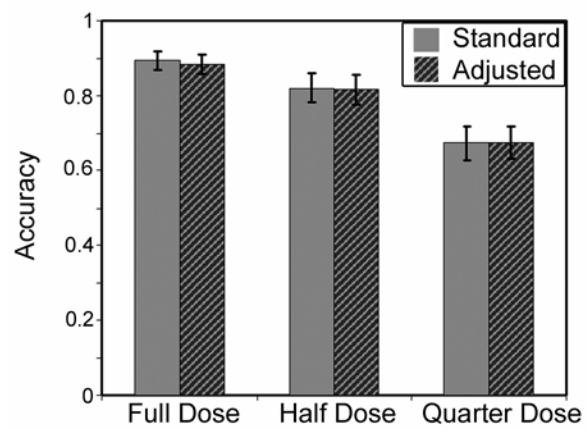Overall performance

Figure 4

Task Performance

Figure 5

Figure 6
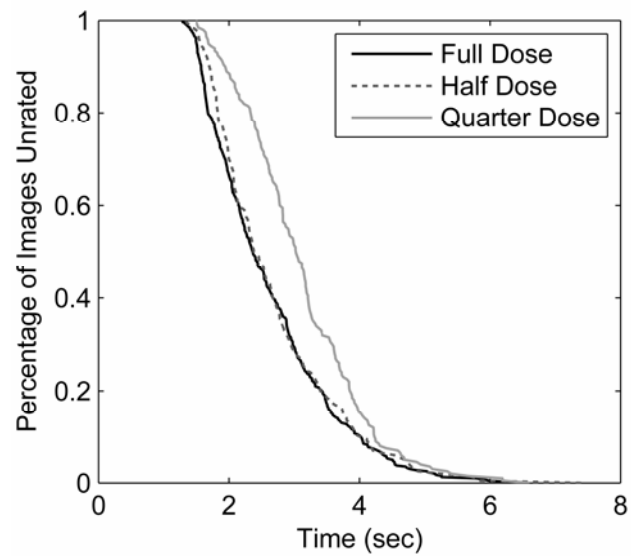

Figure 7

Improving mammographic decision accuracy by incorporating observer ratings with interpretation time

[1]R S SAUNDERS, PhD and [1,2]E SAMEI, PhD

[1]*Duke Advanced Imaging Laboratories, Department of Radiology, 2424 Erwin Road, Suite 302, Duke University Medical Center, Durham, NC 27705 and [2]Departments of Physics, Biomedical Engineering and Medical Physics, Duke University, Durham, NC 27710, USA*

ABSTRACT. Mammography is currently the most established technique for the early detection of breast cancer. However, mammography would benefit from further improvements as it does produce some errors, such as not finding all early-stage cancers. The objectives of this study were first, to measure the timing of correct and incorrect reading decisions in mammography and second, to exploit those dependencies to improve accuracy in mammographic interpretation. To address these objectives, an experiment was conducted where experienced breast imaging radiologists reviewed 400 mammographic regions equally divided among images that contained simulated benign masses, malignant masses, malignant microcalcifications and no lesions. The experiment recorded the radiologists' decision as well as the length of time the mammogram was interpreted in. The experiment results showed that incorrect detection as well as incorrect classification decisions were associated with longer interpretation times ($p<0.0001$). The timing results were used to create a model that would flag cases for review that had a higher probability of error. The flagged cases had a median accuracy drop of 13% for detection decisions and 16% for classification decisions compared with unflagged cases. This suggests that interpretation time can be incorporated into mammographic decision-making in order to identify cases with higher probabilities of perceptual error that require further review.

Mammographic interpretation is a difficult perceptual task, with 20–40% of cancers missed in the initial mammographic screening [1–4]. In addition to missed cancers, another perceptual error is the substantial number of false positives as the specificity of mammography ranges from 88% to 98% [1–3]. Reducing the number of missed cancers and increasing specificity in mammography should be one of the goals of perception science.

Previous perception studies have decomposed interpretation errors into three categories, based on the length of time the radiologist focuses on a potential lesion: search errors, recognition errors and decision-making errors [5, 6]. These studies indicate that search errors occur when the radiologist does not focus once on the abnormality; recognition errors happen when the radiologist briefly examines a potential abnormality, but dismisses it very quickly; and decision-making errors arise when a radiologist examines a potential abnormality for a extended period of time, but still incorrectly classifies it [5]. Some previous studies have investigated these perceptual errors, but have generally considered them together [5, 7–9].

This study focused on the third category, decision-making errors. For screening, these errors occur after the radiologist has searched the image and recognized the area as a potential abnormality, but then incorrectly classifies the area as not containing a lesion. These errors can be more difficult to avoid than other perceptual errors because while improving the conspicuity of lesions can be expected to reduce search errors and recognition errors, it would not necessarily improve decision-making performance. In fact, decision-making errors have been suggested to be the primary perceptual errors in chest radiography [10]. To better understand and decrease decision-making errors, the purpose of this study was two-fold: (1) to measure the timing of correct and incorrect reading decisions in mammography and (2) to exploit those dependencies to improve the accuracy of mammographic interpretation.

Methods and materials

This study isolated decision-making errors by controlling the search process and lesion variability. An image set of 400 mammographic regions was created by inserting simulated breast masses and microcalcifications into digital mammograms. The mammographic

regions were then reviewed by experienced breast imaging radiologists, who rated whether the mammograms contained a lesion or not, and classified the lesion. The ratings and interpretation time for each observer were analysed to understand decision-making errors and whether incorporating interpretation time could improve accuracy.

Mammographic images

A database of 984 de-identified four-view mammograms was obtained with approval by the institutional review board (IRB). Each mammogram had been acquired on an indirect flat-panel mammography detector (GE Senographe 2000D; GE Medical Systems, Waukesha, WI) [11, 12]. Out of this database, 200 craniocaudal views were chosen for further analysis.

Lesion simulation

Simulated mammographic lesions, the realism of which was verified in previous studies, were embedded in the digital mammograms [13–15]. These simulated lesions included typically benign masses (oval circumscribed and oval obscured), typically malignant masses (irregular ill-defined and irregular spiculated), and typically malignant microcalcifications (fine linear branching and clustered pleomorphic). The contrast for these lesions was determined by a Monte Carlo model (xSpect) of the mammographic image acquisition [16]. The contrast was reduced by the expected scatter, which was calculated from previous models [17].

Image processing

The images were processed by a two-stage process to enhance fine detail and provide sufficient contrast at the skin line [18, 19]. After this processing, the histogram of each image was analysed to find the appropriate window and level. The window and level was approximated by a sigmoid curve, which provided a smooth transition at the extremes of the greyscale range. All image processing was evaluated by an experienced breast imaging radiologist (JAB: 7 years experience, 5000 cases per year). The radiologist did not participate in the observer performance experiment to minimize bias.

Observer performance experiment

The mammograms were reviewed by five experienced radiologists (average 11.2 years of experience as radiologist attending, average 9.8 years as mammography attending, average 160 cases per week). The radiologists reviewed images on a custom graphical user interface (GUI) that displayed a 5.12 cm × 5.12 cm region of the mammogram for interpretation. The radiologists rated each image based on whether it appeared to contain microcalcifications, a benign mass, a malignant mass, or no lesion. Images were viewed three times, once on a medical-grade liquid crystal display (LCD) (Nova V; National Display Systems, Morgan Hill, CA; 165 µm pixels) and twice on a medical-grade CRT (MGD 521; Barco LLC, Duluth, GA; 148 µm pixels). For each reading, the interface recorded the radiologist interpretation time, or the interval between the time the mammogram was displayed and the time the radiologist recorded his or her rating.

The observer experiment controlled for other factors by adopting the following constraints. Each image was displayed at full resolution to maintain image fidelity. The image centre was indicated by four whiskers on each side in order to minimize image search. To reduce rating correlations between sequential images, radiologists could not return to an image once it had been rated. The radiologists viewed each display straight ahead and centred as some displays, such as LCDs, have different properties off-axis [20]. To maintain a similar image appearance, the radiologists could not adjust the image window and level. Finally, the display order, image order and image orientation were randomized to further reduce potential biasing effects.

Statistical analysis

The data were analysed to determine the performance at two different clinical tasks. One task was a screening task where radiologists must detect a mammographic lesion. For this detection task, the radiologists would be correct if they detected the lesion, even if they incorrectly classified it as benign or malignant. The other task was a diagnostic task where the radiologists had to differentiate between benign and malignant breast masses. In this classification task, the lesion had been detected and the radiologists were judged on whether the lesion was classified appropriately. For both tasks, accuracy was computed as the average of sensitivity and specificity.

For each task, the data were analysed to learn whether incorrect and correct decisions correlated with different interpretation times. First, the interpretation time was analysed using survival analysis, where the ''survival time'' of an image was defined as the length of time it remained unrated. The survival curves were plotted to qualitatively show whether rating errors affected interpretation time. Next, the interpretation time distributions for correct and incorrect ratings were compared using statistical tests. A Wilcoxon test compared the centre of the distributions, while a Brown-Forsythe test compared the width of the distributions [21]. Finally, the interpretation times were modelled as a function of decision type (*e.g.* true positive, false positive) using a Proportional Hazards model, allowing a further test of whether rating errors affected interpretation time [22, 23].

After testing whether decision types had a statistically significant impact on interpretation time, two models were constructed to exploit that information. The first model used a nominal logistic regression fit to fit the mammogram truth as a function of the observer ratings alone, interpretation time only, or observer ratings combined with interpretation time. This fit was then used to predict mammogram truth for given observer data (either ratings, timing, or ratings plus timing). The

second model operated similarly to a computer aided detection (CAD) system as it did not make decisions on the mammogram truth, but rather flagged cases with higher probability of incorrect decisions for further review by radiologists. To decide which cases to flag, a linear discriminant was used to find a threshold time that best separated false positives from true positives and false negatives from true negatives. Cases with interpretation times above these thresholds were flagged as they had greater probability of being incorrect. These flagged cases should then be given further review by radiologists in order to improve their accuracy. Each model was evaluated for sensitivity, specificity and accuracy with the variance of each quantity estimated using a bootstrap with 10 000 samples [24].

Results

Detection task interpretation time

Figure 1 demonstrates the timing results for the detection task. The figure shows that incorrect decisions had longer interpretation times than true decisions. The interpretation time differences between the four decision categories (false positives, false negatives, true positives and true negatives) were statistically significant both in terms of the mean time (Wilcoxon's $\chi^2=676$, degrees of freedom (DOF)=3, $p<0.0001$) and the timing variance (Brown-Forsythe's $F=78.5$, DOF=3, $p<0.0001$). As shown in Table 1, false positives had statistically significant longer interpretation times than true positives and false negatives had longer interpretation times than true negatives. The interpretation time's correlation with decision category was confirmed with a Proportional Hazards model. This model also found that decision categories had a statistically significant effect on interpretation time ($\chi^2 =462$, DOF=3, $p<0.0001$).

Table 2 illustrates the results of the first predictive model incorporating interpretation time. The table shows that a model based on interpretation time and observer ratings performs slightly better than a model based on observer ratings alone, but not by a statistically significant amount. Interestingly, a model based solely on interpretation time generally performs above chance by a statistically significant amount, suggesting that interpretation time does provide useful information for predicting mammographic truth.

Figure 2 illustrates the results of the second model which flagged suspicious cases for further review. The figure shows that flagged cases generally had statistically significant drops in sensitivity and specificity. Table 3 shows the magnitude of the accuracy drop from the unflagged cases to the flagged cases. For each observer, there was a statistically significant drop in accuracy for the flagged cases.

Classification task interpretation time

Figure 3 illustrates the difference in interpretation times for correct and incorrect classifications of masses. For each observer, incorrect decisions had longer interpretation times. As with detection task, the mean of the interpretation times were different for correct and incorrect decisions (Wilcoxon's $\chi^2=269$, DOF=3, $p<0.0001$) and the width of the interpretation times distributions differed between incorrect and correct classification decisions (Brown-Forsythe's $F=37.1$, DOF=3, $p<0.0001$). The relationship of interpretation time to decision category (false positive, true positive, false negative, true negative) was confirmed using a Proportional Hazards model. This model also found decision categories had a statistically significant effect on interpretation time ($\chi^2 =191$, DOF=3, $p<0.0001$).

Figure 4 shows the results of the flagging model for this classification task. For each observer, sensitivity and specificity dropped for the flagged cases. Table 4 illustrates that there is a statistically significant difference drop in accuracy for flagged cases.
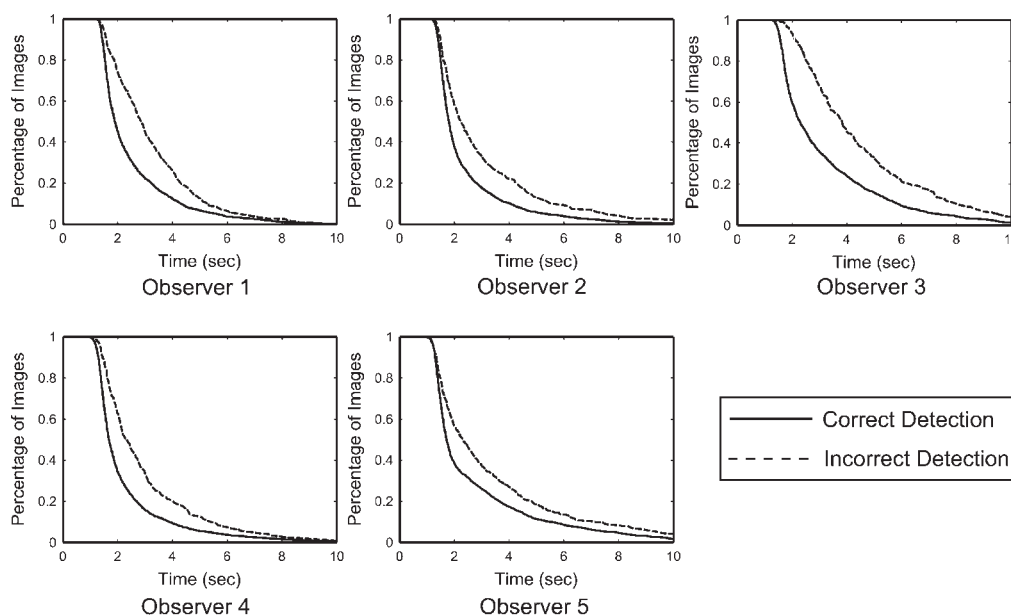

Figure 1. Interpretation times for correct and incorrect detection for detection task.

Table 1. Median interpretation time for different contingency table conditions. The error bars represent the 95% confidence interval of the median

Observer	True negative	False negative	True positive	False positive
1	2.19 + 0.14	2.51 + 0.20	1.84 + 0.05	3.62 + 0.42
2	1.94 + 0.08	2.09 + 0.15	1.76 + 0.03	3.67 + 0.81
3	3.90 + 0.23	4.01 + 0.32	1.99 + 0.05	3.25 + 0.65
4	1.95 + 0.16	2.03 + 0.18	1.64 + 0.03	2.94 + 0.32
5	1.82 + 0.15	2.32 + 0.22	1.70 + 0.04	2.72 + 2.07

Discussion

There has been previous work in investigating perceptual errors. One common means of investigation has been eye-position analysis [5, 7–9]. Eye position analysis infers the type of error based on the amount of time the radiologist focused on a potential abnormality. Eye-tracking relies on the central assumption that foveal attention indicates visual processing of particular areas. This introduces some uncertainty into the results, as foveal focus can include at least a 1° range. Notwithstanding these limitations, previous eye-tracking experiments largely agree with our detection timing results. For pulmonary nodule detection, incorrect decisions were associated with longer interpretation times for experienced radiologists [25]. For breast cancer screening, previous studies found that false positive results from normal mammograms had longer interpretation times than true positive results [5, 8, 9, 26] and false negative results had longer times than true negative results [5].
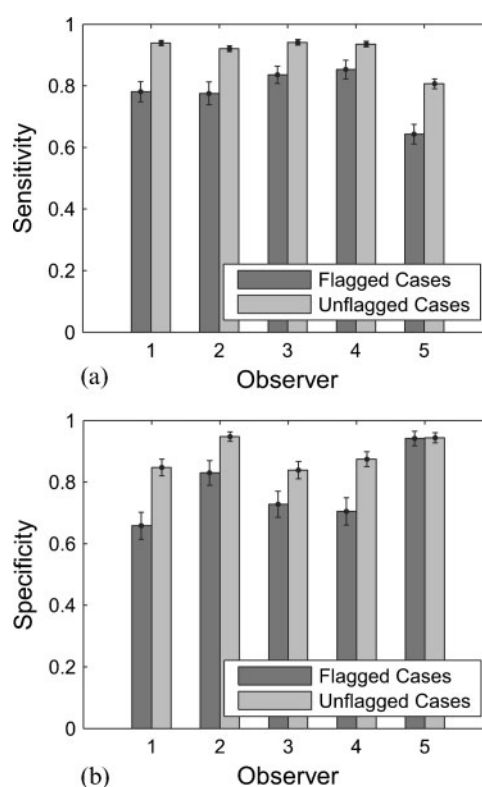
This study showed that interpretation time did correlate with decision category. These results could then be exploited. While a predictive model using interpretation time and observer ratings did not produce statistically significant improvements over a model using observer ratings alone, a flagging model similar to CAD systems did show promise. The flagging model could be used clinically to indicate mammograms requiring further review and potentially improve both the sensitivity and specificity of screening and diagnostic mammography.

In conclusion, this study investigated the potential for using interpretation time as a means of improving accuracy in screening and diagnostic tasks. Detection errors and classification errors had longer interpretation times than correct detection and classification decisions. Using linear discriminant analysis, we established a

flagging program to highlight cases that had a greater probability of incorrect detection or classification decisions. The flagging creates an opportunity to improve mammographic accuracy by identifying cases with statistically lower sensitivities and specificities for further review.

Table 2. Accuracy of models that incorporate rating data only, timing data only, or combine rating and timing data. The error bars represent the 95% confidence interval of the mean

Observer	Accuracy		
	Rating only	Timing only	Rating + timing
1	0.84 + 0.03	0.65 + 0.03	0.86 + 0.02
2	0.91 + 0.02	0.60 + 0.04	0.91 + 0.02
3	0.86 + 0.03	0.76 + 0.03	0.86 + 0.02
4	0.87 + 0.02	0.63 + 0.03	0.88 + 0.02
5	0.85 + 0.02	0.53 + 0.04	0.86 + 0.02

Figure 2. Differences in (a) sensitivity and (b) specificity for detection task with interpretation time flagging.

Table 3. Improvement in detection accuracy of unflagged cases over flagged cases. An asterisk indicates a statistically significant difference. The error bars represent the 95% confidence interval of the mean

Observer	Difference
1	$0.17 \pm 0.06^{*}$
2	$0.13 \pm 0.06^{*}$
3	$0.11 \pm 0.06^{*}$
4	$0.13 \pm 0.06^{*}$
5	$0.08 \pm 0.05^{*}$

Table 4. Improvement in classification accuracy of unflagged cases over flagged cases. An asterisk indicates a statistically significant difference. The error bars represent the 95% confidence interval of the mean

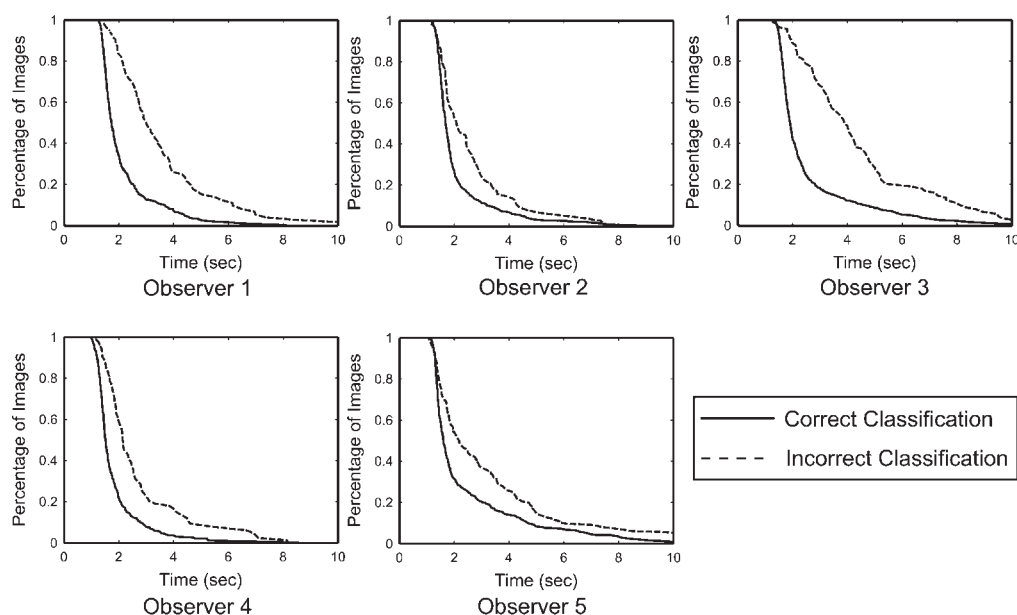| Observer	Difference
1	$0.16 \pm 0.09^{*}$
2	$0.16 \pm 0.09^{*}$
3	$0.19 \pm 0.08^{*}$
4	$0.11 \pm 0.09^{*}$
5	$0.14 \pm 0.09^{*}$

Figure 3. Interpretation times for masses correctly and incorrectly classified as benign or malignant.
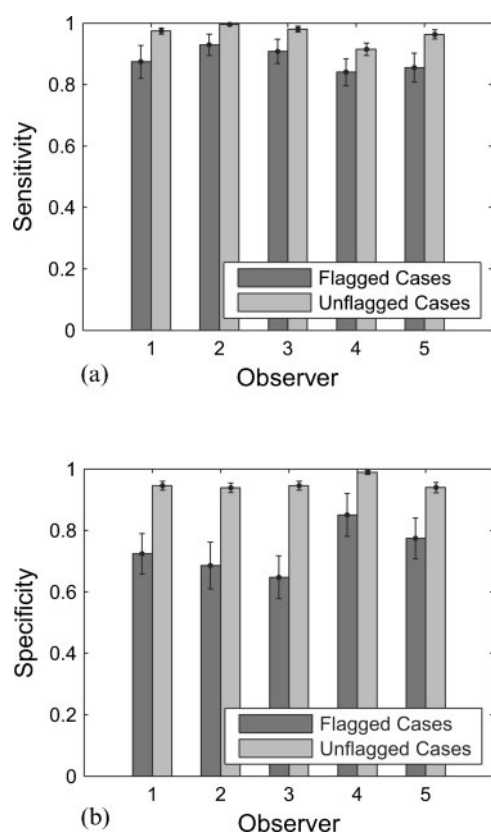
Figure 4. Differences in (a) sensitivity and (b) specificity for classification task with interpretation time flagging.

References

1. Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. Sydney Breast Imaging Accuracy Study: comparative sensitivity and specificity of mammography and sonography in young women with symptoms. AJR Am J Roentgenol 2003;180:935–40.

2. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138:168–75.

3. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353:1773–83.

4. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992;184:613–7.

5. Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP. Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol 2002;179:917–23.

6. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978;13:175–81.

7. Nodine CF, Mello-Thoms C, Weinstein SP, Kundel HL, Conant EF, Heller-Savoy RE, et al. Blinded review of retrospectively visible unreported breast cancers: an eye-position analysis. Radiology 2001;221:122–9.

8. Mello-Thoms C, Hardesty L, Sumkin J, Ganott M, Hakim C, Britton C, et al. Effects of lesion conspicuity on visual search in mammogram reading. Acad Radiol 2005;12:830–40.

9. Mello-Thoms C, Britton C, Abrams G, Hakim C, Shah R, Hardesty L, et al. Head-mounted versus remote eye tracking of radiologists searching for breast cancer: a comparison. Acad Radiol 2006;13:203–9.

10. Manning DJ, Ethell SC, Donovan T. Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. Br J Radiol 2004;77:231–5.

11. Suryanarayanan S, Karellas A, Vedantham S. Physical characteristics of a full-field digital mammography system. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 2004;533:560–70.

12. Vedantham S, Karellas A, Suryanarayanan S, Albagli D, Han S, Tkaczyk EJ, et al. Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype. Med Phys 2000;27:558–67.

13. Saunders Jr RS, Samei E, Baker JA. Simulation of breast lesions. In: 7th International Workshop on Digital Mammography; Durham, NC; 2004.

14. Saunders RS, Samei E. Characterization of breast masses for simulation purposes. Proc SPIE 2004;5372:242–50.

15. Saunders RS Jr, Samei E, Baker JA. Simulation of mammographic lesions. Acad Radiol 2006;13:860–70.

16. Samei E, Flynn MJ. An experimental comparison of detector performance for direct and indirect digital radiography systems. Med Phys 2003;30:608–22.

17. Boone JM, Lindfors KK, Cooper VN 3rd, Seibert JA. Scatter/primary in mammography: comprehensive results. Med Phys 2000;27:2408–16.

18. Stahl M, Aach T, Dippel S. Digital radiography enhancement by nonlinear multiscale processing. Med Phys 2000;27:56–65.

19. Davies AG, Cowen AR, Parkin GJS, Bury RF. Optimizing the processing and presentation of PPCR imaging. Proceedings of SPIE - The International Society for Optical Engineering 1996;2712:189–95.

20. Krupinski EA, Johnson J, Roehrig H, Nafziger J, Lubin J. On-axis and off-axis viewing of images on CRT displays and LCDs: observer performance and vision model predictions. Acad Radiol 2005;12:957–64.

21. Heiberger RM, Holland B. Statistical analysis and data display: an intermediate course with examples in S-plus, R, and SAS. New York, NY: Springer, 2004.

22. Cox DR. Regression models and life-tables. J Royal Statistical Society. Series B (Methodological) 1972;34:187–220.

23. Lawless JF. Statistical models and methods for lifetime data. New York, NY: Wiley, 1982.

24. Efron B, Tibshirani R. An introduction to the bootstrap. New York, NY: Chapman & Hall, 1993.

25. Manning D, Barker-Mill SC, Donovan T, Crawford T. Time-dependent observer errors in pulmonary nodule detection. Br J Radiol 2006;79:342–6.

26. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, et al. How experience and training influence mammography expertise. Acad Radiol 1999;6:575–85.

Comparison of LCD and CRT Displays Based on Efficacy for Digital Mammography[1]

Robert S. Saunders, Ehsan Samei, Jay Baker, David Delong, Mary Scott Soo, Ruth Walsh, Etta Pisano
Cherie M. Kuzmiak, Dag Pavic

Rationale and Objectives. To compare two display technologies, cathode ray tube (CRT) and liquid crystal display (LCD), in terms of diagnostic accuracy for several common clinical tasks in digital mammography.

Materials and Methods. Simulated masses and microcalcifications were inserted into normal digital mammograms to produce an image set of 400 images. Images were viewed on one CRT and one LCD medical-quality display device by five experienced breast-imaging radiologists who rated the images using a categorical rating paradigm. The observer data were analyzed to determine overall classification accuracy, overall lesion detection accuracy, and accuracy for four specific diagnostic tasks: detection of benign masses, malignant masses, and microcalcifications, and discrimination of benign and malignant masses.

Results. Radiologists had similar overall classification accuracy (LCD: 0.83 ± 0.01, CRT: 0.82 ± 0.01) and lesion detection accuracy (LCD: 0.87 ± 0.01, CRT: 0.85 ± 0.01) on both displays. The difference in accuracy between LCD and CRT for the detection of benign masses, malignant masses, and microcalcifications, and discrimination of benign and malignant masses was -0.019 ± 0.009, 0.020 ± 0.008, 0.012 ± 0.013, and 0.0094 ± 0.011, respectively. Overall, the two displays did not exhibit any statistically significant difference ($P > .05$).

Conclusion. This study explored the suitability of two different soft-copy displays for the viewing of mammographic images. It found that LCD and CRT displays offer similar clinical utility for mammographic tasks.

Key Words. Digital mammography; observer performance; soft-copy display; liquid crystal display; LCD; cathode ray tube; CRT.

©AUR, 2006

A recent study has demonstrated the effectiveness of digital mammography in detecting early-stage breast cancer, especially in dense breast tissue (1). The results of that study will encourage the increased use of digital mammography for clinical screening. Digital mammography differs from film-screen systems in that it separates image acquisition, image processing, and display components such that each may be independently optimized (2–4). To optimize the display components, it must be determined how soft-copy displays affect diagnostic performance for specific mammographic tasks, such as the detection of microcalcifications and masses. This question is important both to assess the clinical utility of each display system and to provide data for individual radiology practices when purchasing new display systems.

Two competing soft-copy display technologies are commonly used to display digital images, cathode ray tube (CRT) and liquid crystal display (LCD). Traditionally, soft-copy reading was done on CRTs, but LCDs are rapidly becoming more commonly used for reading medical images. Because these displays rely on different physical processes, their resolution and noise properties differ markedly. CRTs form images by an electron beam striking a phosphor layer, which produces visible light. They exhibit markedly lower resolution than LCDs primarily because of light scattering inside the phosphor layer and the width of the electron beam (5,6). Their resolution further degrades over time from decreasing phosphor efficiency and the necessitated increases in the electron beam intensity (7,8). LCDs use liquid crystal elements to precisely control the amount of light from each pixel and are known for their excellent resolution, which is predominantly limited by their pixel size (9). Therefore, a CRT and LCD with similar nominal pixel sizes and matrix sizes would have different resolution properties because of their pixel structure (2). However, LCDs exhibit significant fixed pattern noise because of the electronics needed to operate the liquid crystals within each pixel (10). In contrast, the noise levels of a CRT, governed primarily by phosphor grain nonuniformities at the faceplate, are often lower than those of the LCDs (2). It is unclear how the differing resolution and noise characteristics of LCDs and CRTs affect the clinical utility of these soft-copy displays.

The purpose of this study was to measure how well breast-imaging radiologists perform clinical tasks using typical LCD and CRT displays. For each display, the overall classification accuracy and overall lesion detection performance was calculated. In addition, this study examined human performance at specific clinical tasks, including the detection of benign masses, malignant masses, and microcalcifications, and the discrimination of benign and malignant masses. By examining human performance on these tasks, we aimed to determine if one technology merits preferential use in digital mammography.

MATERIALS AND METHODS

The study was conducted in multiple steps. First, simulated masses and microcalcifications were inserted into normal digital mammograms using an established simulation routine. The images were rated by experienced breast-imaging radiologists, classifying each image ac-

Table 1
Specific Properties of the LCD and CRT Displays Used in this Study

	CRT	LCD
Manufacturer	Barco, LLC	National Display Systems
Model	MGD 521	Nova V
Display card	Barco MP1H (10-bit)	RealVision MD5mp (10-bit)
Pixel pitch (mm)	0.148	0.165
Matrix size	2048 × 2560	2048 × 2560
Active display area	304 mm × 380 mm	338 mm × 422 mm
L_{min} (cd/m^2)	0.52	0.52
L_{max}(cd/m^2)	308	371

CRT: cathode ray tube; LCD: liquid crystal display.

cording to what type of lesion, if any, it contained. The observer data were analyzed to assess overall classification accuracy and performance on specific clinical tasks. The following sections detail each of these steps.

Display Devices

This study compared two commercial medical displays, an LCD and a CRT (2,11). Table 1 shows the specifications of each display, including information provided by the display manufacturer and luminance measurements conducted in our laboratories (2,11). The two displays were similar in terms of matrix size and pixel pitch. Both displays were calibrated according to the Digital Imaging and Communications in Medicine and American Association of Physicists in Medicine TG18 standards before use (12,13); all other properties, such as luminance, stayed at the default manufacturer setting.

Mammographic Backgrounds

With prior permission from the Institutional Review Board, 200 craniocaudal images were selected from a deidentified database of digital mammograms acquired on a clinical indirect flat panel mammography system (GE Senographe 2000D, GE Medical Systems, Waukesha, WI) (14,15). The images were acquired with a molybdenum anode, molybdenum or rhodium filtration, and a tube potential range from 25 to 30 kVp. The image set had compressed breast thicknesses ranging from 2.7 cm to 7.4 cm and breast compositions ranging from almost entirely fat to extremely dense.

Lesion Simulation

To investigate lesion detection and discrimination, an established lesion simulation routine was used to insert

simulated masses and microcalcifications into the mammograms. This routine relied on the measured characteristics of real lesions to create simulated lesions with a realistic appearance. The routine produced three different types of lesions: typically malignant masses (modeled after *irregular ill-defined* and *irregular spiculated* masses), typically benign masses (modeled after *oval circumscribed* and *oval obscured* masses), and typically malignant microcalcifications (modeled after *fine linear branching* and *clustered pleomorphic* microcalcifications). Breast imaging radiologists have previously confirmed the realism of our simulated lesions (16–18).

The lesion size and subtlety were chosen based on a preliminary experiment to achieve 80% overall classification accuracy. By choosing contrasts that led to this level of overall classification accuracy, the lesions were not so subtle that the observers missed all of them, but not so conspicuous as to be detected by every observer. To achieve this accuracy, masses needed to have a diameter of 3.3–4.1 mm. Although these sizes might be smaller than that typically acted on in the clinic, they provided the appropriate detection level for our "location-known-exactly" study. Individual microcalcifications were 0.35 mm in average diameter inside microcalcification distributions of 4–7 mm diameter, which is similar to that encountered in standard clinical practice. The lesions were scaled to the appropriate contrast as determined by an x-ray model (xSpect software) (19), assuming an average breast (50% glandular/50% adipose tissue) and accounting for anode material, tube filtration, beam energy, and compressed breast thickness. The lesion contrast was further reduced according to the expected scatter to primary ratios computed based on previous investigations (20), accounting for beam energy and compressed breast thickness, with the scatter to primary ratios corrected for the scatter rejection by the antiscatter grid. The scatter to primary ratios for our mammograms, after correction for the antiscatter grid, ranged from 0.07 to 0.22 with an average of 0.14. The simulated lesions were added to the mammograms in a logarithmic scale to model the x-ray attenuation process. The scatter-adjusted logarithmic contrasts for masses averaged 0.069 (ranging from 0.048 to 0.10) and for microcalcifications averaged 0.10 (ranging from 0.070 to 0.14).

Figure 1 illustrates how lesions were paired with normal mammograms to create a set of 400 images. For two tasks, mammographic backgrounds were paired to reduce statistical variance. For example, one mammographic
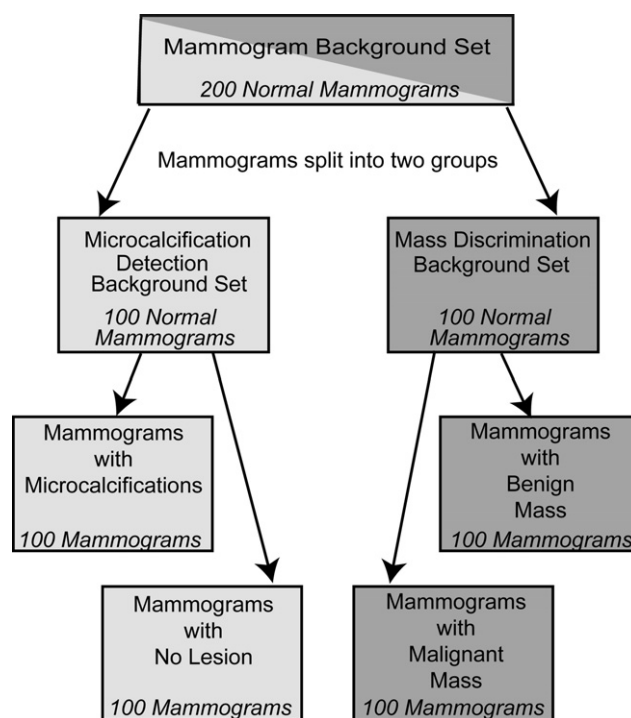

Figure 1.   Distribution of images in reading set.

background produced two images—one with a microcalcification lesion and one with no lesion. This particular scheme was chosen to minimize the number of times a particular mammographic background was viewed by a reader, reducing potential memory problems.

Image Postprocessing

All detector manufacturers apply image postprocessing to the raw detector image to create an image appearance acceptable to radiologists. The lesion simulation routine required raw images from the detector, which prevented the use of manufacturer postprocessing. Therefore, a basic two-stage postprocessing algorithm, boosting fine and broad contrast details (21,22), was used to create image appearances typical of those used clinically. The algorithm applied identical image processing to all images to eliminate the confounding effects associated with variations in image postprocessing. After this application, a histogram analysis was used to determine an optimal window and level setting for each image. The determined window and level was fit by a sigmoid curve to provide a smooth transition at the extremes of the grayscale range and the sigmoid function was then applied to all images. The appropriateness of the window and level settings was verified by a radiologist with 7 years experience in breast

imaging, reading about 5000 cases/year (JAB). This radiologist did not participate in the later observer experiment to minimize bias.

Observer Performance Experiment

Five experienced breast imaging radiologists participated in the observer experiment, representing two different academic medical centers. The radiologists had an average reading volume of 160 cases per week (ranging from 80 to 300) of screening mammography. They had served an average of 11.2 years as a radiology attending (range: 6–17 years) with an average of 9.8 years as an attending in mammography (range: 3–17 years).

The radiologists reviewed the images in a room with low ambient lighting using a customized graphic user interface. This interface was developed to emulate the clinical paradigm while minimizing reading time. The interface displayed a 5.12 cm × 5.12 cm image extracted from the center of each mammogram (Fig. 2). Images were shown sequentially, one at a time. All images were displayed at full resolution (one image pixel represented by one display pixel). After viewing the displayed mammogram, the radiologist rated the image into one of four categories: microcalcifications present, a benign mass present, a malignant mass present, or no lesion present. The radiologists were asked to view each display straight ahead and centered to minimize any confounding effects from off-axis viewing (23). The radiologists were allowed to choose their viewing distance based on their comfort with most choosing a distance of approximately 50 cm. To maintain the consistency of the image appearance for all observers, observers were not allowed to window and level the images.

The experiment began with a training set of 100 images to familiarize the radiologist with the lesion types and the graphic user interface, with feedback given to the radiologist after each image was rated. This proceeded to the reading set, consisting of 2 sessions of 200 images viewed on each of the two display devices for a total of 800 ratings (2 sessions × 200 images × 2 displays). This rating scheme improved statistical power as it controlled for image effects because we could compare an image's rating when it was viewed on a LCD versus when it was viewed on an a CRT. To minimize potential biasing effects, the display order, the image order, and the image orientation were randomized and radiologists were given a 5-minute break between sessions.

Statistical Analysis

The observer data were first analyzed to show how images with a given lesion were rated. These ratings were summarized in contingency tables in which each element, labeled $\zeta_{Truth, Rating}$, represented the number of images from a given truth state that were rated into a given rating category. The contingency tables were further summarized into an overall classification accuracy metric representing the percentage of mammograms correctly rated by an observer as

$$\text{Overall Classification Accuracy} = \frac{\zeta_{NN} + \zeta_{CC} + \zeta_{BB} + \zeta_{MM}}{\text{Total Number of Cases}},$$

(1)

where $\zeta_{Truth, Rating}$ represents the number of images from a given truth state that were rated into a given rating category, N corresponds to no lesion category, C to the microcalcification category, B to benign mass category, and M to malignant mass category. The associated variance was calculated using a bootstrap analysis, which resampled the image set into 10,000 bootstrap samples (24). The overall classification accuracy and its associated variance were calculated both individually and jointly for each display and for each observer. The contingency tables were also summarized into a metric examining overall lesion detection accuracy, which was computed as the average of sensitivity and specificity of detecting a lesion. For overall lesion detection accuracy, a true positive was defined as detecting a lesion within an abnormal mammogram, even if the observer misclassified the lesion as benign or malignant. The variance for overall lesion detection accuracy was also calculated using bootstrap analysis.

The data were also analyzed for accuracy at several clinical tasks, including the detection of microcalcifications, detection of benign masses, detection of malignant masses, and the discrimination of benign and malignant masses. For the example task of microcalcification detection, sensitivity and specificity were calculated as follows:

$$\begin{aligned} &Sensitivity: \frac{\zeta_{CC}}{\zeta_{CC} + \zeta_{CN}}\\ &Specificity: \frac{\zeta_{NN}}{\zeta_{NN} + \zeta_{NC}} \end{aligned}.$$

(2)

where $\zeta_{Truth, Rating}$ represents the number of images from a given truth state that were rated into a given rating cate-
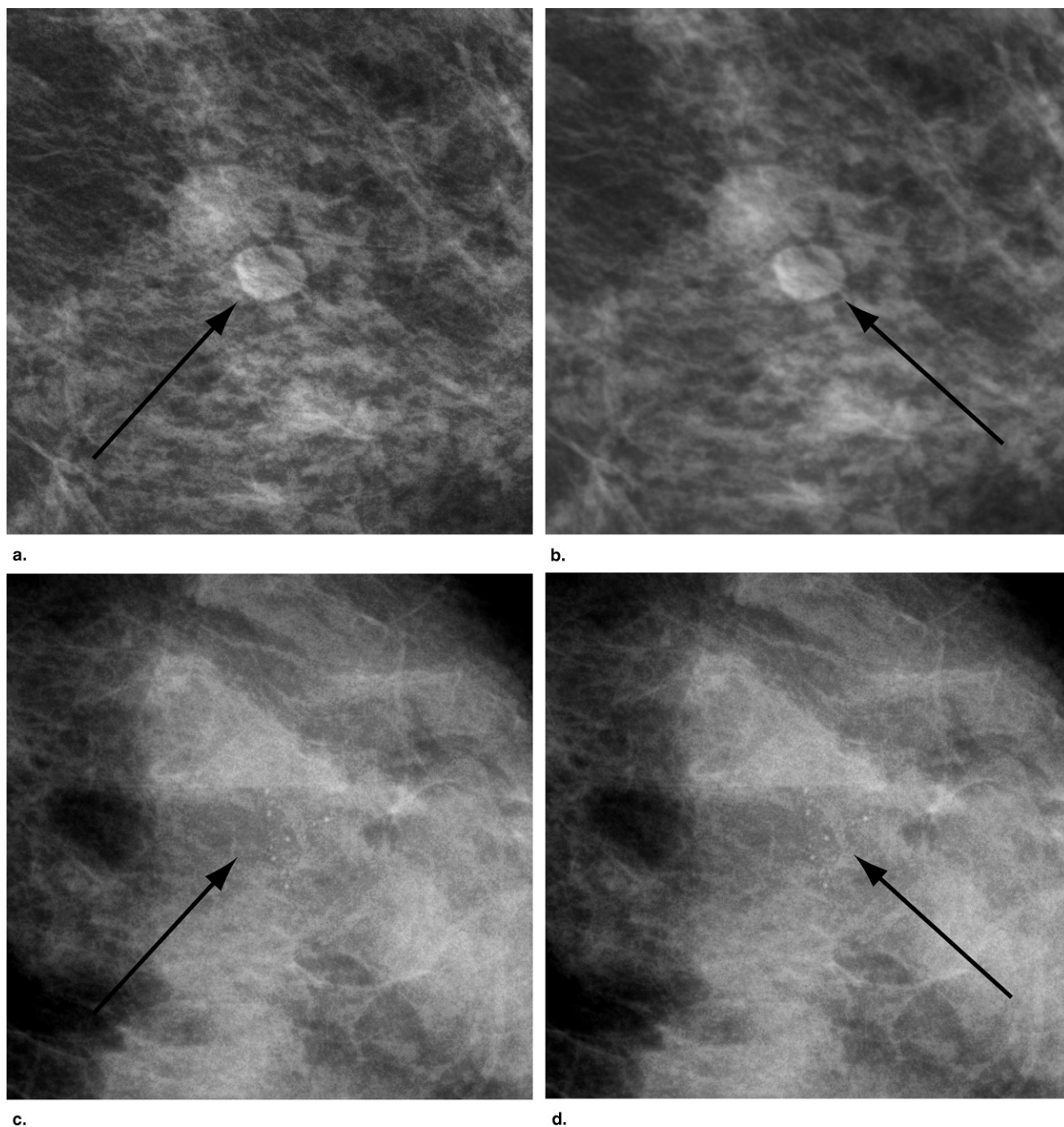
Figure 2. Example images at resolutions corresponding to the liquid crystal display (LCD) and cathode ray tube (CRT) displays. The rows include benign masses **(a, b)** and microcalcifications **(c, d)**, with the left column shows images at LCD resolution and the right column shows images at CRT resolution.

gory, *N* corresponds to no lesion category, and *C* to the microcalcification category. The average of sensitivity and specificity was the task accuracy. The associated variances for each clinical task were similarly calculated both individually and jointly for each observer and for

each display using bootstrap analysis, which resampled the paired mammographic backgrounds into 10,000 bootstrap samples. For each task, statistical significance was estimated using a *P* values generated by a paired *t*-test (25).

Table 2
Mean Performance on Two Displays, Including Overall Classification Accuracy, Overall Lesion Detection Accuracy, and Task Performance for Four Clinical Tasks, Averaged Over All Observers

	LCD	CRT	P Value of Difference
Overall classification accuracy	0.83 ± 0.01	0.82 ± 0.01	.63
Overall lesion detection	0.87 ± 0.01	0.85 ± 0.01	.22
Detection of microcalcifications	0.89 ± 0.02	0.91 ± 0.01	.28
Detection of benign masses	0.96 ± 0.01	0.94 ± 0.01	.23
Detection of malignant masses	0.88 ± 0.02	0.87 ± 0.02	.52
Discrimination of benign and malignant masses	0.93 ± 0.01	0.92 ± 0.01	.55

CRT: cathode ray tube; LCD: liquid crystal display.

Table 3
Average Observer Ratings for Each Image Truth State

	Rating			
LCD	Normal	Microcalcification	Benign Mass	Malignant Mass
Truth				
Normal	87%	2.6%	2.2%	8.2%
Microcalcification	18%	78%	0.8%	4.0%
Benign mass	5.4%	0.0%	88%	7.0%
Malignant mass	15%	1.0%	5.4%	79%

	Rating			
CRT	Normal	Microcalcification	Benign Mass	Malignant Mass
Truth				
Normal	83%	3.8%	2.2%	11%
Microcalcification	13%	84%	0.6%	2.0%
Benign mass	8.8%	0.4%	82%	9.2%
Malignant mass	14%	1.2%	4.8%	80%

CRT: cathode ray tube; LCD: liquid crystal display.

RESULTS

In terms of overall classification accuracy, the LCD and CRT appeared to offer similar performance (P = .63), as shown in Table 2. This similarity also held true for overall lesion detection, which measured the ability of observers to detect a lesion even if they misclassified it as benign or malignant. Note that overall classification accuracy would theoretically range from 0.25 (chance in a four-category scheme) to 1 (perfect accuracy), whereas the detection accuracies would range from 0.5 (chance) to 1 (perfect accuracy). Table 2 also shows performance metrics for four specific clinical tasks, along with associated variances, averaged across all observers. All four specific task performances appeared similar for both displays with all differences well within the associated vari-

ances of the experiment. Although the accuracy tended to be higher for the LCD for most tasks, the CRT had higher accuracy for the microcalcification detection. None of these differences was statistically significant ($P \geq .23$).

Table 3 shows the contingency tables for the average observer. The contingency tables show that most images were classified correctly on both displays. As shown in Fig. 3, an examination of the misclassifications reveals that benign masses were sometimes rated as malignant masses and vice versa, indicating observers had some difficulties discriminating the two mass types. However, the contingency tables for the CRT and LCD were generally similar, implying that the observers performed similarly on both displays.

Performance averaged across observers might mask differences at the individual observer level. Figure 4 illus-
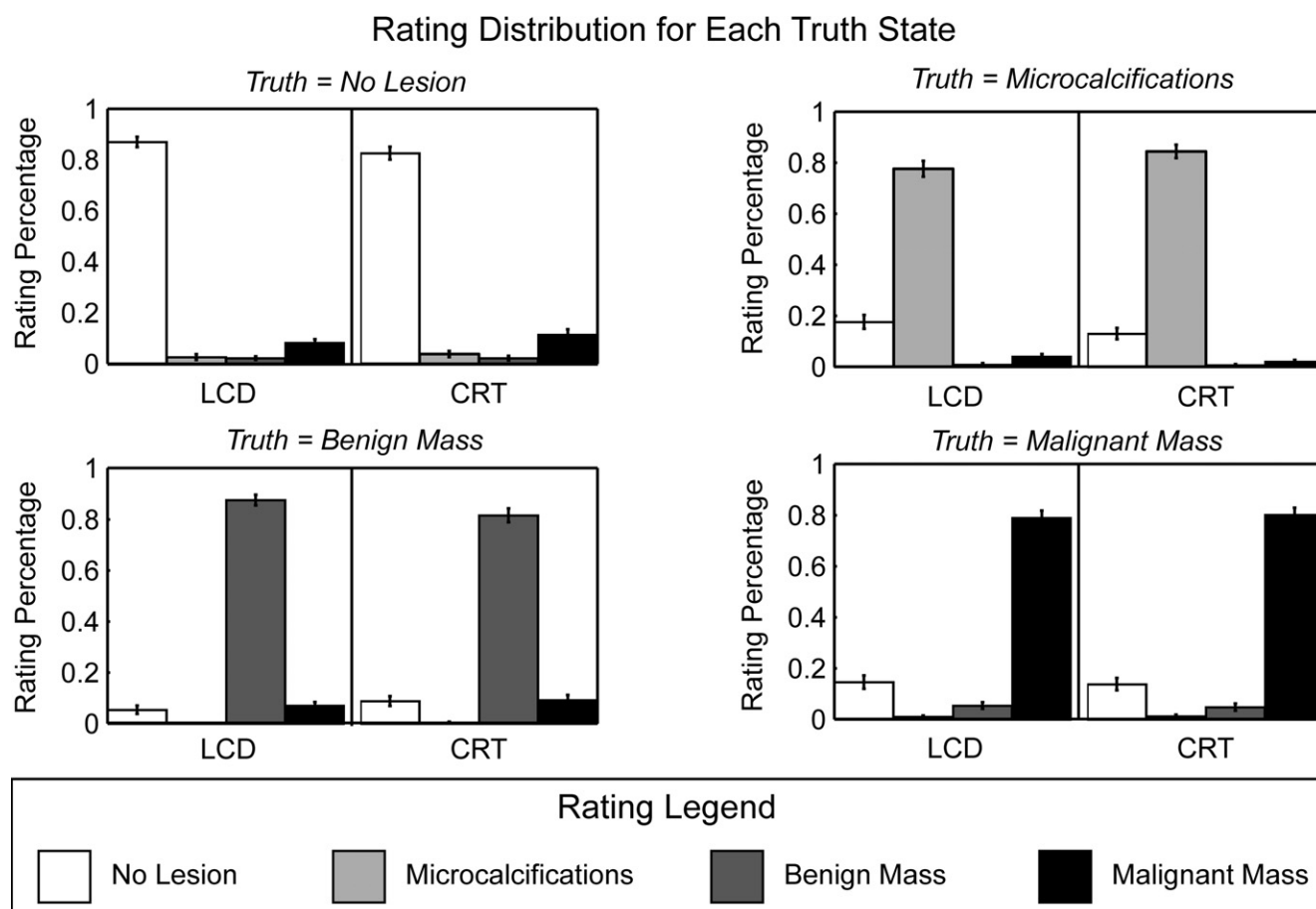
Rating Distribution for Each Truth State

Figure 3. Rating distribution for each truth state and each display for the average observer.
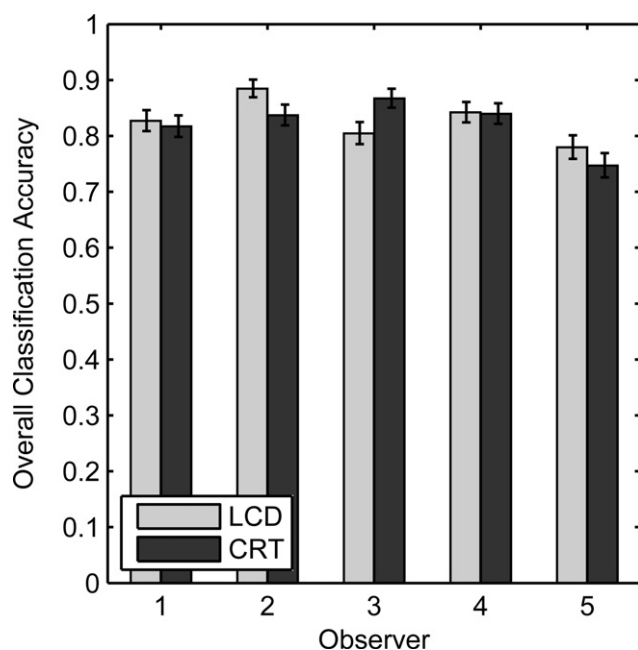

**Figure 4.** Overall classification accuracy for each observer.

trates the overall classification accuracy for each observer and display. Observers A and D had similar performance on each display, observers B and E achieved slightly higher accuracy on the LCD, whereas observer C showed the opposite effect. This trend among observers generally held true for the individual task performances, except for the detection of microcalcifications where observers A, C, and D performed slightly better on the CRT than on the LCD. The individual observer results suggest that some observers might perform better on particular display devices, although there is not sufficient statistical power to substantiate this claim.

DISCUSSION

This study examined how different digital displays affect clinical performance in mammography. By inserting simulated masses and microcalcifications into normal clinical cases, this investigation could control for lesion
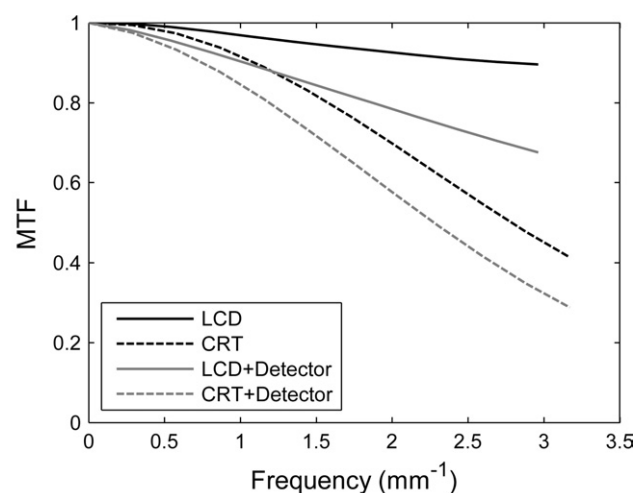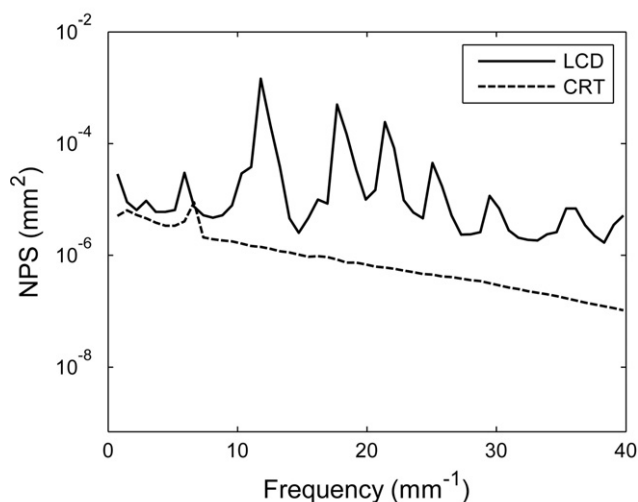
Figure 5. Resolution (left, **a**) and noise (right, **b**) evaluated in this experiment. The resolution graphs plot the resolution of the displays and the resolution of the entire system, accounting for both display and detector. The detector resolution, reported previously (14), had been appropriately adjusted to account for the display magnification by scaling the frequencies by the ratio of the display pixel size to the detector pixel size assuming each detector pixel is represented by a single display pixel.

size, contrast, breast density, and anatomic features. Lesion detection and discrimination were determined through an observer performance experiment employing five experienced radiologists. The results of this study suggested that display modality (LCD vs. CRT) had little impact on diagnostic accuracy.

The two displays yielded similar performance, even though the resolution of the two displays differed substantially, as shown in Fig. 5a (2). This similarity in performance may be explained on several grounds. First, the

resolution advantages of the LCD may be offset by its increased noise, as shown in Fig. 5b (2). Second, the similar performance may have resulted from the limited resolution of the mammography detector (15). To account for this influence, Fig. 5a plots the resolution of the entire imaging system, including the resolution of the display devices and the resolution of the detector. The difference in resolution between the two displays was reduced after accounting for the resolution of the detector. However, the LCD detector system still has superior resolution to the CRT detector system. Third, although the displays have different resolution and noise, the limitations of the human visual system may mean that human observers perceive the two displays to have similar resolution and noise properties (2). Finally, the similarity may be explained by the substantial experience of the radiologists deployed in our study. All radiologists in this study were experienced in breast imaging and read a substantial number of cases per year. Radiologists with less experience may be more affected by different displays. Regardless of the reasons, our study clearly found that experienced radiologists had similar accuracy at clinical tasks on the LCD and CRT displays.

There has been little prior work examining the impact of different soft-copy display devices on clinical performance. Three earlier works examined observer performance for different chest radiography tasks, the detection of pulmonary nodules and detection of catheters, but did not find statistically significant differences between LCDs and CRTs (26–28). A study on breast mass detection (23,29) found that LCDs yielded slightly better performance than CRTs, but not by a statistically significant margin ($A_z = 0.91 \pm 0.01$ for LCDs vs. $A_z = 0.90 \pm 0.02$ for CRTs) (23). Our study examined a wider range of clinical tasks, including the detection of benign and malignant masses and the detection of microcalcifications and the discrimination of masses. In addition, our work employed more than twice the number of anatomical backgrounds (200 mammographic backgrounds versus 80 in other work) and mass templates (200 versus 80). The findings of our study are consistent with that of previous studies in concluding that the impact of display modality on diagnostic accuracy is extremely limited (23).

This experiment differed from previous studies in that it used a categorical rating paradigm. This type of scoring improved the throughput of the observer experiments, allowing observers to view the 800 images of this study in a short time. The gains in throughput may lower the variance, as the observer could rate more images in a

shorter period of time although on a coarser scale. In addition, the rating paradigm emulated the clinical situation closely, because clinical situations often demand binary decisions about the presence or state of an abnormality. Most other competing methods, such as receiver operating characteristic analysis, are based on confidence ratings and therefore differ from clinical paradigm.

Clinical interpretations of mammograms require decisions that are almost entirely binary in nature. In each clinical mammogram interpretation, radiologists include a final assessment category from the Breast Imaging Reporting and Data System (BI-RADS) which is a 7-point scale from 0 to 6 (30). Although the BI-RADS final assessment category provides different levels of confidence for the presence of breast cancer, only specific categories can be used in different settings, making the decisions largely binary. For example, when interpreting screening mammograms, the interpreting physician determines whether the mammogram warrants further evaluation (BI-RADS category 0) or does not (BI-RADS category 1 or 2). Categories 3–6 are rarely used in the screening setting. In the diagnostic setting, after all mammogram and ultrasound images have been reviewed, the decision is again largely binary based on whether the lesion in question warrants a biopsy (category 4 or 5) or is it definitively benign (category 1 or 2) based on imaging alone. Whether a lesion's BI-RADS final assessment is category 4 or category 5 has no impact on the ultimate recommendation to biopsy the lesion, because virtually all such lesions will undergo tissue sampling. Although a lesion could also be interpreted as BI-RADS 3, probably benign, in the diagnostic setting, this category is infrequently employed when used as intended by the BI-RADS manual (30). In a recent study by Kerlikowske et al, 1.6% of patients undergoing their first screening mammograms and only 0.7% of subsequent screening exams were assigned BI-RADS final assessment category 3 after an appropriate diagnostic evaluation (31). Therefore, the clinical decision in breast imaging can be summarized as answering the binary question "further evaluation needed or not" for a screening exam or "biopsy recommended or not" for a diagnostic study.

This study faced certain limitations. First, the lesion contrast was calculated for an average breast (50% glandular/50% adipose tissue). This was an approximation, because the mammogram database contained images of breasts with various compositions. Second, all mammograms for this study were acquired on an indirect flat-panel detector. Mammograms obtained on other digital detectors might appear slightly different on the LCD or CRT displays. Third, the radiologists knew whether they were using an LCD or CRT, adding a potential source of bias to the experiment. Fourth, this study only used specific, though typical, LCD and CRT displays. Other display devices may offer slightly different performance. Fifth, we used simulated breast masses and microcalcifications to create abnormal mammograms. Although breast-imaging radiologists had previously confirmed the realism of our simulated lesions (16–18), simulated lesions may not represent all of the natural variability of breast lesions. This highlights the importance of a randomized clinical trial to confirm the clinical performance of each display device. Finally, we displayed all images at full resolution on each display. Some display workstations may use an alternative display protocol in which the images are not displayed at full resolution. Images displayed at reduced or enlarged sizes will experience blurring from the display and by the interpolation algorithm used by the display card and display software. This may result in slightly different performance for the radiologists in those clinical settings. Notwithstanding these limitations, we believe that the study provides a reasonable evaluation of the effects of current display technologies on mammographic task performances.

In summary, this work explored the impact of different soft-copy displays on common mammographic tasks. By using simulated masses and microcalcifications, the study controlled for contrast, size, and breast background. The results indicate that CRTs and LCDs yield similar performance, even though observer dependent performance trends cannot be ruled out. Although the resolution of the two displays differs markedly, it appears that resolution may be only one of many factors impacting the clinical utility of a display. These results are particularly relevant for radiology practices evaluating the costs and benefits of different display systems for clinical breast imaging.

REFERENCES

1. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353:1773–1783.
2. Saunders RS, Samei E. Resolution and noise measurements of five CRT and LCD medical displays. Med Phys 2006; 33:308–319.
3. Saunders RS Jr, Samei E, Jesneck JL, et al. Physical characterization of a prototype selenium-based full field digital mammography detector. Med Phys 2005; 32:588–599.
4. Fan J, Dallas WJ, Roehrig H, et al. Improving visualization of digital mammograms on the CRT display system. Proc SPIE 2003; 5029:746–753.

5. Mertelmeier T. Why and how is soft copy reading possible in clinical practice? J Dig Imaging 1999; 12:3–11.

6. Muka E, Blume HR, Daly SJ. Display of medical images on CRT soft-copy displays: a tutorial. Proc SPIE 1995; 2431:341–359.

7. Compton KD. Factors affecting CRT display performance: specifying what works. Proc SPIE 2000; 3976:412–423.

8. Samei E, Flynn MJ. A method for in-field evaluation of the modulation transfer function of electronic display devices. Proc SPIE 2001; 4319: 599–607.

9. Blume HR, Steven PM, Cobb ME, et al. Characterization of high-resolution liquid crystal displays for medical images. Proc SPIE 2002; 4681: 271–292.

10. Badano A, Gagne RM, Jennings RJ, et al. Noise in flat-panel displays with subpixel structure. Med Phys 2004; 31:715–723.

11. Samei E, Wright SL. Luminance and contrast performance of liquid crystal displays for mammographic applications. Technol Cancer Res Treatment 2004; 3:429–436.

12. NEMA. Digital Imaging and Communications in Medicine (DICOM) Part 14: grayscale display standard function. Rosslyn, Va: National Electrical Manufacturers Association, 2000.

13. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. Med Phys 2005; 32:1205–1225.

14. Suryanarayanan S, Karellas A, Vedantham S. Physical characteristics of a full-field digital mammography system. Nucl Instrum Methods 2004; 533:560–570.

15. Vedantham S, Karellas A, Suryanarayanan S, et al. Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype. Med Phys 2000; 27:558–567.

16. Saunders RS Jr, Samei E, Baker JA. Simulation of breast lesions. In: 7th International Workshop on Digital Mammography. Durham, NC, 2004; 162–169.

17. Saunders RS, Samei E. Characterization of breast masses for simulation purposes. Proc SPIE 2004; 5372:242–250.

18. Saunders RS Jr, Samei E, Baker JA, et al. Simulation of mammographic lesions. Acad Radiol 2006; 13:860–870.

19. Samei E, Flynn MJ. An experimental comparison of detector performance for direct and indirect digital radiography systems. Med Phys 2003; 30:608–622.

20. Boone JM, Lindfors KK, Cooper VN 3rd, et al. Scatter/primary in mammography: comprehensive results. Med Phys 2000; 27:2408–2416.

21. Stahl M, Aach T, Dippel S. Digital radiography enhancement by nonlinear multiscale processing. Med Phys 2000; 27:56–65.

22. Davies AG, Cowen AR, Parkin GJS, et al. Optimizing the processing and presentation of PPCR imaging. Proc SPIE 1996; 2712:189–195.

23. Krupinski EA, Johnson J, Roehrig H, et al. On-axis and off-axis viewing of images on CRT displays and LCDs: observer performance and vision model predictions. Acad Radiol 2005; 12:957–964.

24. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall, 1993.

25. Heiberger RM, Holland B. Statistical analysis and data display: an intermediate course with examples in S-plus, R, and SAS. New York: Springer, 2004.

26. Hwang SA, Seo JB, Choi BK, et al. Liquid-crystal display monitors and cathode-ray tube monitors: a comparison of observer performance in the detection of small solitary pulmonary nodules. Korean J Radiol 2003; 4:153–156.

27. Oschatz E, Prokop M, Scharitzer M, et al. Comparison of liquid crystal versus cathode ray tube display for the detection of simulated chest lesions. Eur Radiol 2005; 15:1472–1476.

28. Scharitzer M, Prokop M, Weber M, et al. Detectability of catheters on bedside chest radiographs: comparison between liquid crystal display and high-resolution cathode-ray tube monitors. Radiology 2005; 234: 611–616.

29. Krupinski EA, Johnson J, Roehrig H, et al. Use of a human visual system model to predict observer performance with CRT vs LCD display of images. J Dig Imaging 2004; 17:258–263.

30. D'Orsi C. Illustrated Breast Imaging Reporting and Data System (BI-RADS). Reston, Va: American College of Radiology, 1998.

31. Kerlikowske K, Smith-Bindman R, Abraham LA, et al. Breast cancer yield for screening mammographic examinations with recommendation for short-interval follow-up. Radiology 2005; 234:684–692.

Simulation of Mammographic Lesions[1]

Robert Saunders, Ehsan Samei, Jay Baker, David Delong

Rationale and Objectives. This study presents a method for generating breast masses and microcalcifications in mammography via simulation. This simulation method allows for the creation of large image datasets with particular lesions, which may serve as a useful tool for perception studies measuring imaging system performance.

Materials and Methods. The study first characterized the radiographic appearance of both masses and microcalcifications, examining the following five properties: contrast, edge gradient profile of masses, edge characteristics of masses, shapes of individual microcalcifications, and shapes of microcalcification distributions. The characterization results then guided the development of routines that created simulated masses and microcalcifications. The quality of the simulations was verified by experienced breast imaging radiologists who evaluated simulated and real lesions and rated whether a given lesion had a realistic appearance.

Results. The radiologists rated real and simulated lesions to have similarly realistic appearances. Using receiver operating characteristic analysis to characterize the degree of similarity, the results showed an A_z of 0.68 ± 0.07 for benign masses, 0.65 ± 0.07 for malignant masses, and 0.62 ± 0.07 for microcalcifications, thus showing notable overlap in the simulated and real lesion ratings.

Conclusion. This research introduced a new approach for simulating breast masses and microcalcifications that relied on anatomic characteristics measured from real lesions. Results from an observer performance experiment indicate that our simulation routine produced realistic simulations of masses and microcalcifications as judged by expert radiologists.

Key Words. Simulation; mammography; lesion modeling; observer performance.

© AUR, 2006

Several studies have advocated evaluating an imaging system using task-based diagnostic approaches (1–5). These approaches, instead of relying on the measurement of image quality solely based on physical metrics, examine how well a physician performs a clinical task using images from a given imaging system. For mammography, the performance of a system is measured based on how well it aids the clinician in detecting breast cancer. Measuring cancer detection generally involves human observer performance experiments, in which an observer reads a large number of cases, rating each image based on whether it appears to contain a lesion. Such experiments rely on the availability of a large number of images with a particular lesion class, the presence of which should be confirmed independently. Given the extremely small percentage of cancer cases in the mammography screening population (6), obtaining a large enough database of images with confirmed lesions is no trivial task. Simulation techniques thus present an attractive alternative for investigating lesion detection and classification questions because they allow one to easily form large databases and to investigate large numbers of influencing variables more efficiently.

The current state of the art in simulating breast lesions is relatively limited. Previous investigations have mostly used Gaussian profiles, disks, or simulated lung nodule profiles (7–12) to simulate masses. These prior geometric models are generally overly simplistic to provide an adequately complex representation of real breast masses, which exhibit notable variations from patient to patient in terms of border characteristics, shape, and contrast profile. More realistic approaches have been used to simulate microcalcifications (13–16). However, those methods have mostly relied on templates formed from actual cases, limiting the number of simulations available.

In this study, we developed a new technique to simulate mammographic lesions. First, we measured the physical characteristics of breast masses and microcalcifications in mammographic images. These characteristics guided the development of simulation routines capable of creating breast lesions with realistic and variable appearance. The lesion appearance was finally validated by experienced breast imaging radiologists.

MATERIALS AND METHODS

Lesion Characterization

Characterization of breast masses.—From descriptors in the Breast Imaging Reporting and Data System (BI-RADS) lexicon, four common categories of breast masses were chosen (17). Those categories included two types of typically benign masses (oval circumscribed and oval obscured masses) and two types of typically malignant masses (irregular ill-defined and irregular spiculated). A total of 152 mammograms, each containing a mass described by the one of the above four categories, were drawn from the University of South Florida's Digital Database for Screening Mammography (DDSM) (18). The mammograms were segmented into 2.56 cm × 2.56 cm regions of interest (ROIs) surrounding the mass. For each ROI, the original optical density values were determined using the measured characteristic curve of the scanner reported on the DDSM web site. Each ROI was then analyzed to determine the properties of the masses.

The masses were characterized using a three-stage process: 1) segmenting the mass from the surrounding anatomy, 2) examining the mass contrast profile in terms of an edge gradient profile, and 3) measuring the edge properties in terms of a border deviation profile. The following outlines the steps for each stage in the characterization process. In the first stage, the masses were segmented

from the surrounding anatomy using a Laplacian of Gaussian edge detection method (19). Each segmentation was visually inspected to ensure that it covered the whole mass. The edge outlines were used to create a mask for each mass, which was then fit with an ellipse to determine the major axis length, b_0; the ratio of the minor to major axis length, c; center location (x_0, y_0); and the orientation angle between the major axis and x-axis, α.

In the second stage, the mass contrast profile was measured using an edge gradient profile, formed by averaging pixel values along elliptical rings with the same center, orientation angle, and minor to major axis length ratio as measured earlier (Fig 1a). The edge gradient profile contained minimal contribution from background anatomy, as the background anatomy generally did not exhibit elliptical symmetry and therefore averaged to zero along a ring. The edge gradient profile was then fit with a modified sigmoid curve as

$$f(x) = y_o - (\lambda \gamma b_0 x + \beta)\left[1 - \frac{1}{1 + e^{-(x - \gamma b_0)/(\rho b_0)}}\right], \quad (1.1)$$

where x represents the major axis of the elliptical rings and other parameters are defined in Table 1. The edge gradient profile fit parameters for each mass type were averaged across individual masses to obtain average lesion behavior and further minimize any contribution from background anatomy.

Although the edge gradient profile captured the contrast profile of the mass over its transition from the lesion to the background, it assumed that the mass possessed perfect elliptical symmetry. A real mass would have deviations from an ellipse, as shown in Fig 1b. To capture these features, in the third stage, the relative deviation of the mass mask from its corresponding best-fit ellipse was recorded in a border deviation profile. The border deviation profile resembled a random variate with random phase and was thus summarized by its variance and normalized power spectrum. The variance and power spectrum were averaged over all masses of a certain type to establish average lesion behavior.

Figures 2a and 3a show typical edge gradient profiles for benign and malignant masses, respectively. Edge gradient profiles for most benign masses could be characterized by a distinct mass region followed by a sharp transition to background. Malignant mass profiles differed by exhibiting a more gradual transition to background. This difference between benign and malignant masses was
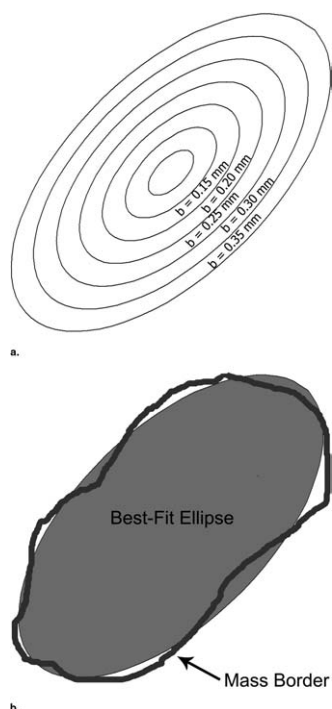
Figure 1. Mass characterization procedure. Illustration of concentric elliptical rings used for the formation of the edge gradient trace with the major axis *b* of the ellipses clearly marked **(a)** and example of border deviation profile **(b)**, which indicates how a mass border differs from a perfect ellipse.

seen in the average profile parameters, listed in Table 1, where the parameter ρ indicated the sharpness (slope) of the transition to background, with smaller values signifying sharper transitions. The benign masses had an average ρ that was approximately half of the average value for malignant masses, indicating benign masses had more distinct borders compared with malignant masses.

Figures 2b and 3b illustrate typical border deviation profiles for benign and malignant masses, respectively. Figures 2c and 3c demonstrate the power spectra from these profiles. Overall, malignant masses showed more deviation from a perfect ellipse than benign masses. This was consistent with the expected behavior of the malignant mass categories as having irregular shapes, whereas the benign mass categories are generally referred to having an oval shape.

Characterization of microcalcifications.—Using descriptors in the BI-RADS lexicon, two common categories of microcalcifications were studied, *fine linear branching* and *clustered pleomorphic*, both representing typically malignant lesions. To study the characteristics of these lesions, 94 mammograms with these types of microcalcifications were drawn from the DDSM. The mammo-

grams were then segmented into 2.56 cm $\times$ 2.56 cm ROIs containing the microcalcification distributions. The ROI pixel values were converted to optical density using the characteristic curve of the scanner. Each ROI was then analyzed to determine the properties of the microcalcifications.

The characteristics of the microcalcifications were evaluated in a three-stage process: 1) segmenting the microcalcifications from the mammogram, 2) measuring the properties of individual microcalcifications, and 3) examining the microcalcification distribution. In the first stage, the lesions were segmented from the background anatomy by thresholding and then manually inspected to ensure all individual microcalcifications were included. In the second stage, individual microcalcification properties were measured from this mask including the major axis length, the minor axis length, and the average contrast. In the third stage, the distribution of individual microcalcifications was determined for clustered pleomorphic and fine linear branching cases. For clustered pleomorphic microcalcifications, the microcalcifications were found to be distributed relatively uniformly within an elliptical area characterized by a major axis and a minor axis for the cluster. The major and minor axis lengths of the cluster were thus recorded. For fine linear branching cases, the microcalcifications were distributed along lines and branches according to the underlying duct structure. Therefore, the distribution properties were characterized in terms of the length of these lines and the angular separation between the lines and branches. Table 2 summarizes the results of the microcalcification characterization.

Lesion Simulation

Mass simulation.—A simulation routine was developed to emulate benign and malignant masses with properties similar to the four chosen categories (benign: oval circumscribed and oval obscured; malignant: irregular ill-defined and irregular spiculated). The mass simulation routine was based on the measured mass characteristics described in the previous section. Using these properties, the simulation routine, as illustrated in Fig 4, consisted of three stages: 1) creation of an array with elliptical rings radiating out from the center; 2) modification of the initial array to produce the proper border shapes; and 3) conversion of the modified elliptical array into pixel values using the edge gradient profile function.

The first stage of the simulation created two arrays that established the elliptical behavior of the masses.

Table 1
Average Parameter Values for the Elliptical Trace for Both Benign and Malignant Masses

Parameter	Symbol	Benign Lesion Average Parameters	Malignant Lesion Average Parameters
Mean background signal	y_0	1.58 ± 0.38	1.63 ± 0.21
Contrast profile in mass region	λ	$-7.39\text{E-}06 \pm 1.80\text{E-}05$	$4.67\text{E-}05 \pm 8.91\text{E-}05$
Mass contrast	β	0.26 ± 0.09	0.48 ± 0.17
Edge location	γ	0.90 ± 0.03	0.49 ± 0.21
Sharpness of edge transition	ρ	0.092 ± 0.018	0.22 ± 0.12
Mass size	b	$5.09 \text{ mm} \pm 1.60 \text{ mm}$	$5.05 \text{ mm} \pm 1.68 \text{ mm}$

The parameter values were determined by fitting the measured edge gradient trace to the curve from Equation 1.1. The parameter ρ determines the sharpness of the mass border, with smaller values indicating a sharper border.

The first array created elliptical rings in which each element in the ring was set equal to the ellipse's major axis value as

$$B_{xy} = \sqrt{\left(\frac{1}{c^2}\right)\Big[\text{Cos}\,[\alpha](x-x_0) - \text{Sin}\,[\alpha](y-y_0)\Big]^2 + \Big[\text{Sin}\,[\alpha](x-x_0) + \text{Cos}\,[\alpha](y-y_0)\Big]^2}, \quad (1.2)$$

where α describes the orientation of the ellipse, c refers to the ratio of minor axis to major axis, and (x_0, y_0) correspond to the center of the ellipse. The second array set each element equal to its angle along the elliptical ring as

$$\Phi_{xy} = \text{ArcTan}\left[\frac{c((y-y_0)\,\text{Cos}\,[\alpha] - (x-x_0)\,\text{Sin}\,[\alpha])}{(x-x_0)\,\text{Cos}\,[\alpha] + (y-y_0)\,\text{Sin}\,[\alpha]}\right], \quad (1.3)$$

with similar parameters as before. For both arrays, the simulation parameters for mass orientation, α, and minor axis to major axis ratio, c, were chosen to match measurements from real lesions.

The second stage of mass simulation established the correct border behavior. To begin, the border was represented as a one-dimensional array of Gaussian random noise, $\zeta(\Phi)$. The noise was then transformed by a fast-Fourier transform and filtered by the measured normalized power spectrum corresponding to the type of mass being simulated. This filtered spectrum was then transformed back to the spatial domain and scaled by the variance of the border deviation profile corresponding to the type of mass being simulated. The deviation profile, $\zeta(\Phi)$, was applied to the elliptical array, B_{xy}, in a multiplicative fashion as

$$B'_{xy} = B_{xy} \cdot (1 + \zeta(\Phi_{xy})). \quad (1.4)$$

In the final stage, the elliptical array was converted to pixel values using the measured edge gradient profile. The simulated mass was then subtracted from a normal mammographic background, N, to produce an output image, O, as

$$O_{xy} = N_{xy} - C \cdot E(B'_{xy}), \quad (1.5)$$

where E represents the normalized edge gradient profile and C corresponds to the lesion contrast.

The contrast of a mass is an important parameter defining its appearance. Because our study was primarily based on screen-film images, the mass contrast was determined by examining the contrast of comparably size masses imaged with identical x-ray tubes and embedded in similarly sized breasts to take into account the nonlinear characteristic curve of such systems. For implementation of our routines for digital systems, a different approach can be taken based on modeling the contrast of a homogenous breast mass, accounting for breast thickness, beam energy, anode target, tube filtration, detector material, and scatter (20,21).

Microcalcification simulation.—A microcalcification simulation routine, summarized in Fig 5, was developed to create simulated microcalcifications based on the clustered pleomorphic and fine linear branching categories. First, this procedure established the microcalcification distribution for clustered pleomorphic and fine linear branching categories. For the clustered pleomorphic case, the microcalcifications were distributed with uniform probability inside an ellipse with major and minor axis lengths as calculated from real cases. Conversely, for the fine linear branching case, the microcalcifications were distributed along lines and branches with lengths and angular separations measured from real cases. After deter-
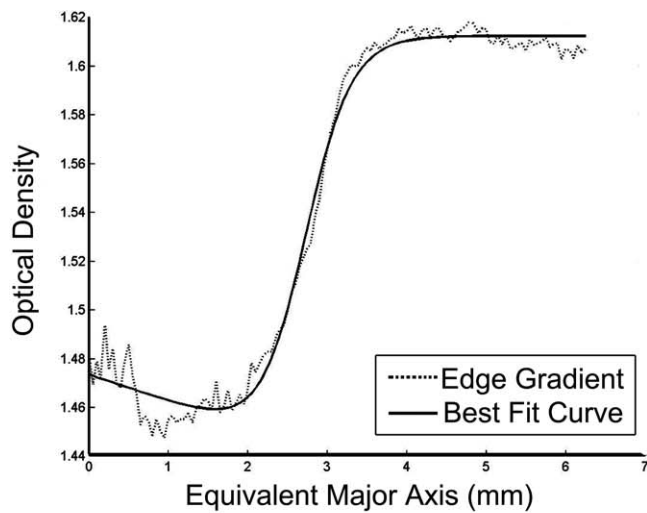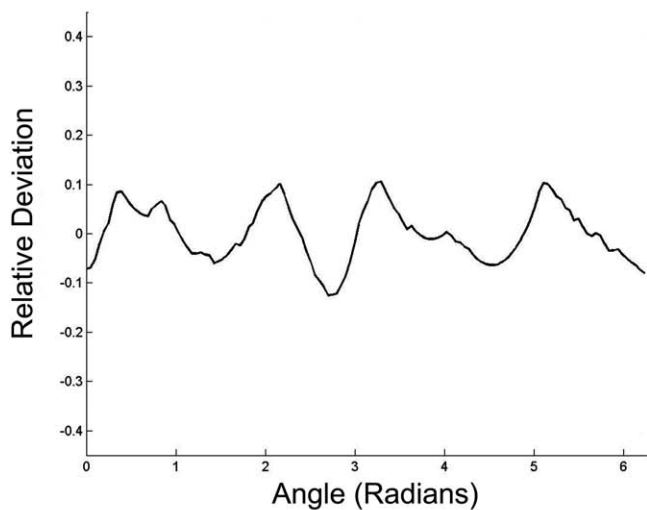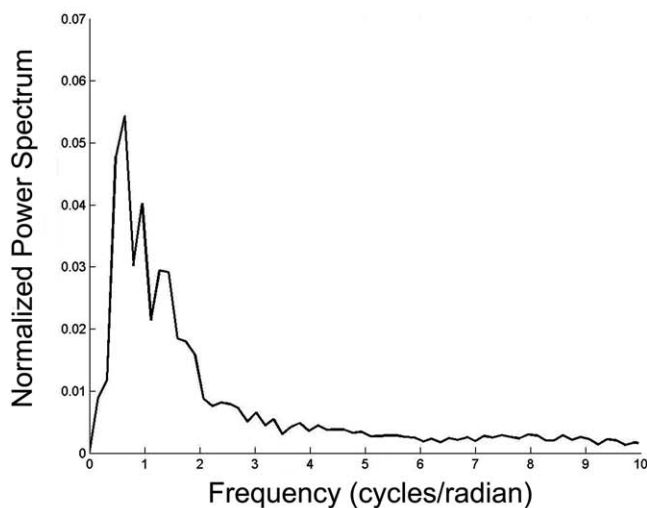
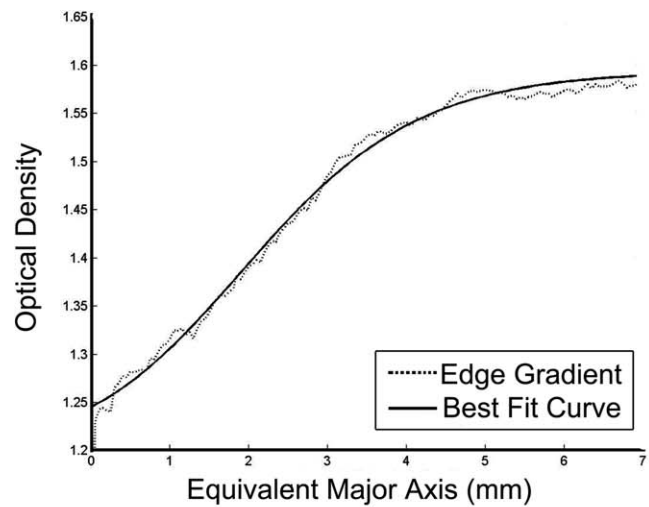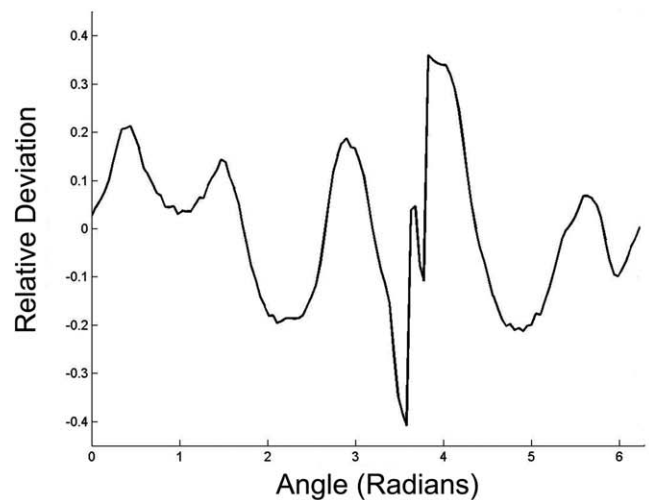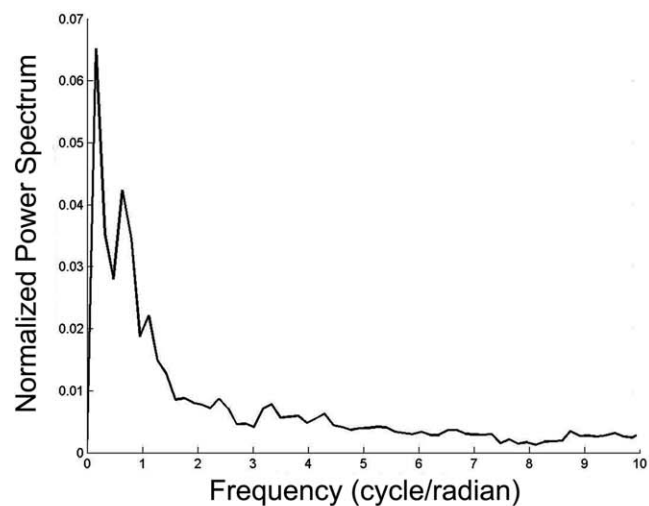**Figure 2.** Characterization results for a typical benign mass: its edge gradient profile **(a)**, border deviation profile **(b)**, and the power spectrum of the border deviation profile **(c)**.

Figure 3. Characterization results for a typical malignant mass: its edge gradient profile **(a)**, border deviation profile **(b)**, and the power spectrum of the border deviation profile **(c)**.

Table 2
Summary of Microcalcification Characterization Results, Including Properties of the Distribution and Individual Microcalcifications

Individual	Pleomorphic	Fine Linear Branching
Microcalcifications		
Major axis (mm)	0.47 ± 0.11	0.43 ± 0.13
Minor axis (mm)	0.29 ± 0.06	0.26 ± 0.05
Contrast	0.22 ± 0.13	0.34 ± 0.16
Distribution		
Major axis (mm)	8.0 ± 3.5	NA
Minor axis (mm)	7.1 ± 3.2	NA
Line length (mm)	NA	6.2 ± 2.3
Angle (degrees)	NA	50.8 ± 11.2

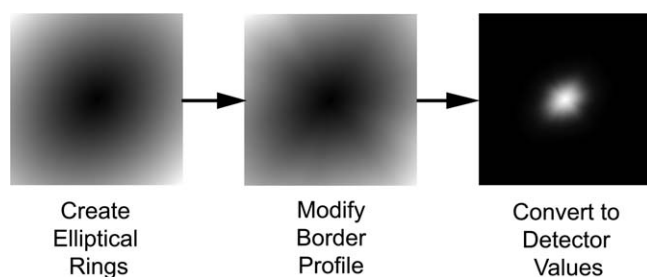Create Elliptical Rings → Modify Border Profile → Convert to Detector Values

Figure 4. Flow chart of mass simulation procedure. The procedure first created elliptical rings radiating outward. The border of each ellipse was modified according to the measured border deviation profile. The simulated mass image was formed by transforming these elliptical rings according to the overall edge gradient.

mining the distribution, the procedure created individual microcalcifications by drawing a line through the microcalcification center at a random angle. The length of this line equaled the major axis length of the individual microcalcifications as calculated from actual microcalcification cases. This line was modified by a morphologic thickening and eroding operation to create realistic edges for individual microcalcifications. Once created, the simulated microcalcifications were then added to a normal background with a given contrast. The exact contrast values were estimated using a procedure similar to that used for masses noted previously.

Observer Performance Experiment

The simulation routine was considered effective if a breast imaging radiologist would judge simulated lesions to have a similarly realistic appearance to that of real ones. To test this hypothesis, an observer performance

experiment was conducted using 200 images containing approximately equal numbers of simulated benign masses, real benign masses, simulated malignant masses, real malignant masses, simulated microcalcifications, and real microcalcifications. Simulated benign masses ranged in size from 4.5 mm to 7 mm, whereas simulated malignant masses had a diameter of 5 mm to 7 mm at their largest extent, and individual microcalcifications measured 250 μm and were located inside distributions measuring 5 mm to 8 mm. The images were viewed on a soft-copy display using a custom graphical user interface. To minimize the effects of display blur, the images were displayed with one image pixel for each display pixel. Three experienced radiologists, with an average of 8 years of breast imaging experience, rated the images on a 100-point scale, where 0 represented "definitely simulated" appearance and 100 represented "definitely real" appearance. The rating experiment placed no constraints on viewing time and the radiologists were allowed to window and level the images as desired. The rating experiment was conducted in a darkened room on a mammographic quality monitor (Barco MGD521, p45 phosphor; BarcoView, LLC; Duluth, GA) calibrated to the DICOM Grayscale Display Function and TG18 standards (22,23).

The rating scores were analyzed using four different methods. The first method compared the ratings for real and simulated lesions qualitatively. Box plots were used to show the similarity in rating distributions for each lesion class and for each observer. Next, the ratings for each image were averaged across observers to find the behavior of the average observer. The average observer's rating distributions were then analyzed in two ways to compare whether the centers of the distributions were different for real and simulated lesions and whether the width of the distributions were different. The difference in the centers of the real and simulated rating distributions was evaluated via a Wilcoxon test, using the normal approximation to find a P value for statistical significance (24) for the second analysis method. For the third, the widths of the distributions were compared using a Brown-Forsythe test (24). Fourth, because a difference may be statistically significant but not practically significant, the data were further analyzed to quantify the degree of overlap between the ratings for real and simulated lesions. This was accomplished using receiver operating characteristic (ROC) analysis. The ROC curves were generated by ROCKIT software (C. Metz, University of Chicago), which also computed the area under the ROC curves, A_z (25). A_z equals 1.0 in a detection experiment when all
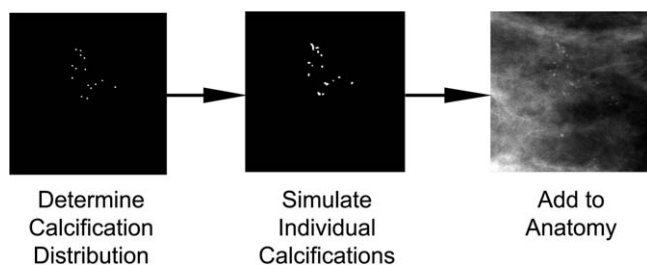
Figure 5. Flow chart of microcalcification simulation procedure. The routine first established the microcalcification distribution, then drew the individual microcalcifications, and finally added the simulated lesion to a normal background with a given contrast.

lesions are detected and equals 0.5 when the observer has chance detection. In contrast to detection experiments, in this analysis, A_z would equal 1.0 when the real and simulated lesions are rated completely differently and 0.5 when real and simulated lesions were rated in exactly the same manner.

RESULTS

Figure 6 shows several examples of simulated benign masses embedded in different mammographic backgrounds. For comparison, real lesions are also shown in this figure. As evident in the figure, the real and simulated lesions share many visual characteristics, as desired. In an analogous fashion, Fig 7 illustrates simulated and real malignant masses in various mammographic backgrounds. Again, the real and simulated masses have similar radiographic appearances. Figure 8 illustrates examples of simulated microcalcification clusters. For comparison, typical real lesions of the same type are also shown. As with masses, the real and simulated lesions possess similar appearances.

Figure 9 illustrates the observer scores for each lesion class. Most notably, the observers gave both real and simulated masses high realism scores. In addition, the observers generally rated simulated lesions similarly to real lesions. The one exception to that observation was Observer 2, who rated real and simulated benign masses slightly differently. However, even for that observer, the simulated benign masses received a high realism score. For malignant masses and microcalcifications, this observer's real and simulated ratings overlapped considerably. Figure 10 shows box plots of the rating distributions for the average observer. The benign mass distributions for the average observer had a statistically significant difference between
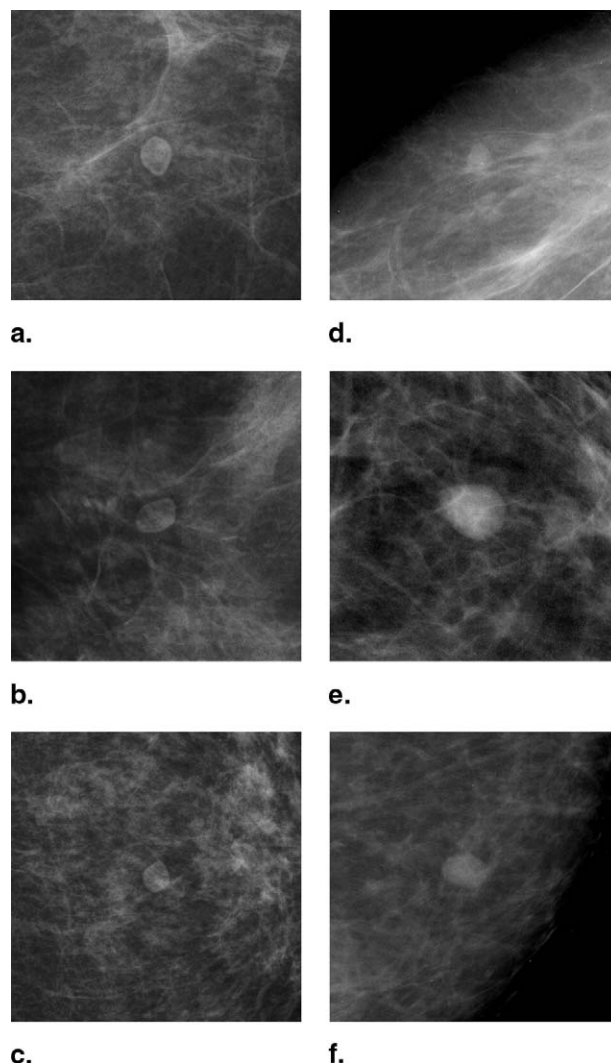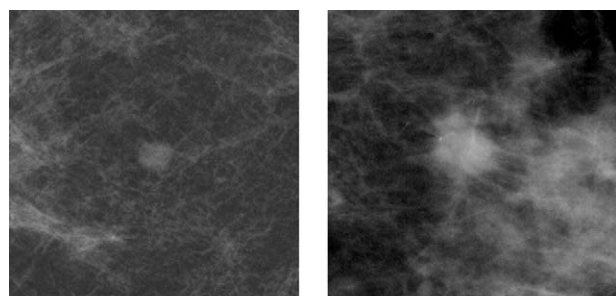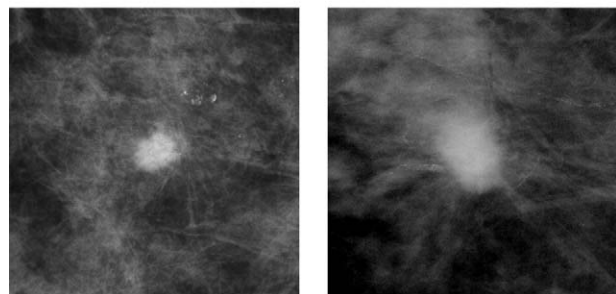

Figure 6. Examples of benign simulated masses embedded in a normal background (left). The figure includes real masses for comparison purposes (right).

real and simulated lesions (Wilcoxon's $z = -2.77$, $P = .0055$), whereas malignant masses and microcalcifications did not exhibit statistically significant differences between real and simulated cases (Wilcoxon's $z = -1.60$, $P = .11$; Wilcoxon's $z = 1.89$, $P = .059$, respectively). The realism ratings had similar variances for real and simulated lesions within each lesion class (benign masses: Brown-Forsythe $F = 2.37$, $P = .13$; malignant masses: Brown-Forsythe $F = 0.96$, $P = .33$; microcalcifications: Brown-Forsythe $F = 2.33$, $P = .13$). Because a statistically significant difference does not indicate a practically significant difference, as noted in the Methods section, the data were further analyzed using ROC analysis to quantify the degree of overlap between real and simulated lesion ratings. All
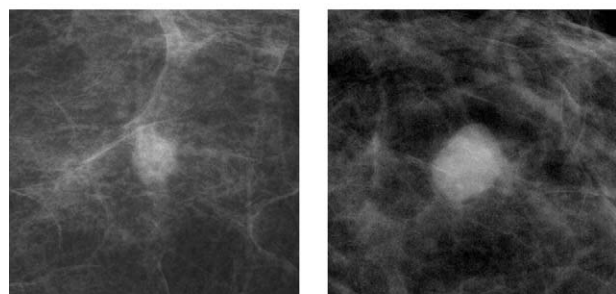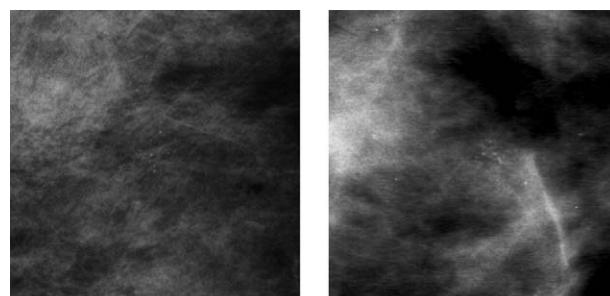
Figure 7. Examples of malignant simulated masses embedded in a normal background (left). The figure includes real masses for comparison purposes (right).

Figure 8. Example simulated microcalcifications (left) along with real lesions (right) for comparison purposes.

lesions have A_z values approaching 0.5, as shown in Table 3, indicating the observers rated simulated lesions to have a similarly realistic appearance to real lesions. In addition, Table 3 shows the absolute difference between real and simulated lesion ratings for each lesion class.
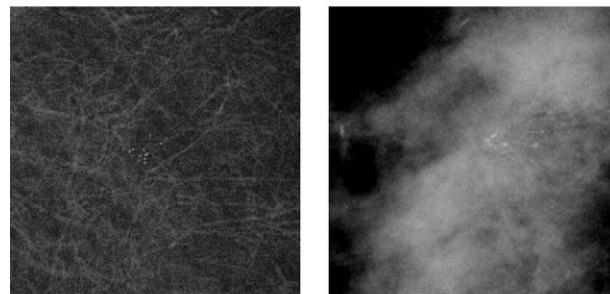
DISCUSSION

The effectiveness of an imaging system lies in its ability to aid in clinical tasks. Because many diagnostic tasks involve detecting lesions, lesion detection experiments are often used to assess the performance of imaging systems.
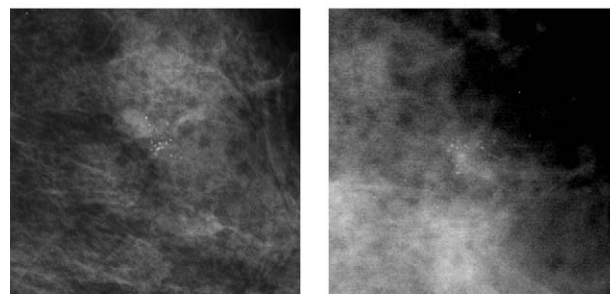
However, these studies are often hindered by the limited number of clinical cases available. This work sought to remedy this problem in mammographic imaging by introducing a new means of simulating breast masses and microcalcifications, enabling the creation of a large number of realistic lesions for detection and discrimination studies.

Most prior attempts at simulating breast masses have used relatively simple approaches. Some prior approaches have relied on Gaussians, blurred discs, or simulated lung nodules to emulate masses (7–12). These simple symmetrical models, however, do not adequately replicate the asymmetrical, complex, and variable appearance of mammographic masses. Another prior approach has been based on templates from masses digitally excised from images (26–29). Although this approach can provide ade-

a.
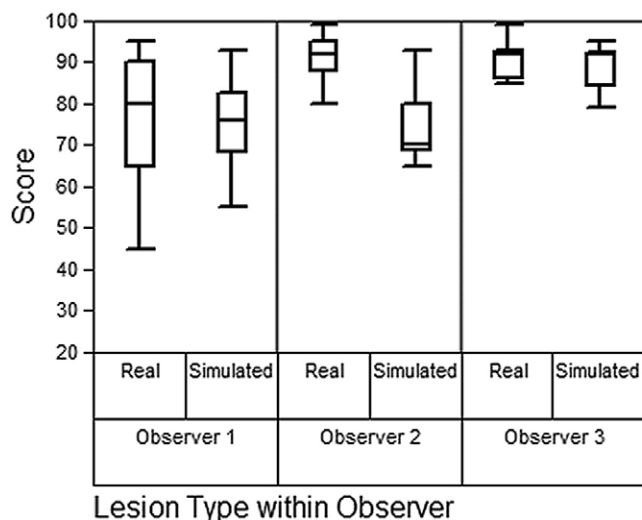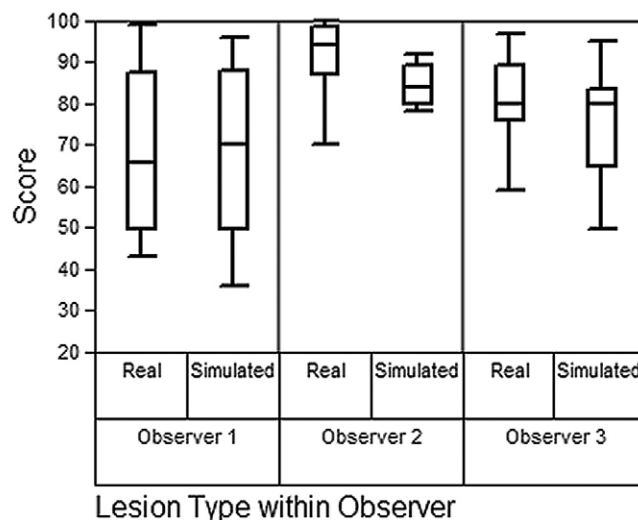
b.
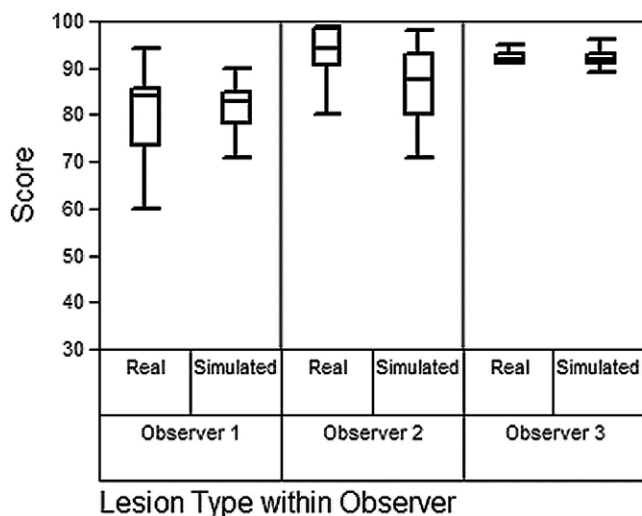
c.

Figure 9. Observer results for benign masses (a), malignant masses (b), and microcalcifications (c). The box plots mark the median as a line within the box, with the top and bottom edges of the box showing the 25th and 75th percentile, respectively. The top and bottom whiskers show the range of the scores (excluding any outliers in the data). For all three lesion classes, the simulated lesions were generally rated to have a similar appearance to real cases.

quate complexity, cases are limited to a handful of lesion templates. This limitation presents a serious problem for investigations that need hundreds or thousands of images for observer or modeling experiments. In addition, this approach does not allow for the control of mass size, which is an important parameter in lesion detectability experiments. A different prior approach has used an anthropomorphic breast phantom to simulate the parenchymal anatomy and lesions (30). This approach was also limited in its ability to generate adequate variability in mass cases for lesion detectability experiments. To over-

come many of the limitations presented by prior approaches, our technique uses measurements from actual mammographic masses to generate simulated masses with a realistic appearance. The appearance of the simulated masses is varied by the routines, enabling the generation of large datasets.

Efforts have also been made to simulate microcalcifications. As with masses, one common prior approach has been the use of blurred discs (10). This geometric model does not capture the complexity of real microcalcifications. Other prior approaches have been based on micro-
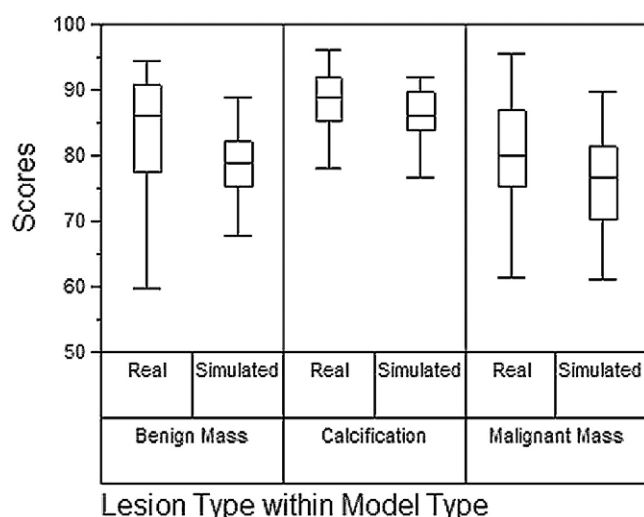
Figure 10. Average observer results for benign masses, malignant masses, and microcalcifications. The box plots show the similarity in realism scores between real and simulated lesions.

Table 3
Summary Statistics for Average Observer

	Difference in Mean Realism Scores (100-point scale)	A_z
Benign masses	6.0 ± 2.2	0.68 ± 0.07
Malignant masses	3.9 ± 2.1	0.65 ± 0.07
Microcalcifications	1.5 ± 1.3	0.62 ± 0.07

The first data column shows the difference between real and simulated realism scores for benign masses, malignant masses, and microcalcifications. The second column quantifies the overlap in realism scores between real and simulated lesions using receiver operator characteristic analysis.

calcification templates digitally removed from actual mammograms (13,14). Again, this template method remains limited by the number of segmented templates available. Another prior approach has been the automatic creation of distributions of microcalcifications, while relying on real templates for individual microcalcifications (15,16). That approach has been restricted to clustered distributions and offers a limited number of templates for individual microcalcifications (15,16). Another prior approach has posited a novel way to make individual microcalcifications, but has not included means to generate a distribution (27). Yet another promising method has used three-dimensional models of microcalcifications, which does not easily extend to existing databases of two-dimensional mammograms (31). To address the limitations of previous methods, our technique relies on the measured

characteristics of real microcalcifications and can generate large numbers of lesions with variable appearance. The current method does not rely on existing templates, allowing for a greater number of images with simulated microcalcifications with either clustered or linear distributions.

This study formed lesion models based on mammograms drawn from the publicly accessible DDSM, which relied on digitized versions of screen-film mammograms. As digital mammography has gained momentum, it could be advantageous to base a lesion-simulation model on digital mammograms. However, at the time of this study, there were not any publicly accessible databases of digital mammograms comparable to the DDSM. The DDSM carries several advantages as well; it contains mammograms with many different lesion types, ranging from benign to malignant masses and multiple types of microcalcifications. The resolution of screen-film systems was high and captured accurate images of the breast anatomy. In terms of contrast, we relied on values based on physical properties that eliminate the impact of film gamma on the lesion contrast. Therefore, although this model is based on film-screen mammograms, it should easily translate to digital mammography.

In this study, we used a ROC methodology to address the quality of the simulated lesions. Another way to verify the realism of the simulated lesions would be a 2-Alternative Forced Choice (2AFC) experiment. However, a 2AFC experiment addresses a slightly different question than the one explored by this study, namely the discriminability of real versus simulated lesions as opposed to assessing the realistic appearance of simulated lesions. Although powerful, a 2AFC experiment can potentially rely on irrelevant details about the simulated lesions to discriminate real from simulated cases. For example, the observer might discern that simulated benign masses have similar shapes, whereas real benign masses have greater shape variability. The observer will likely be able to better discriminate between real and simulated benign masses in this scenario in a 2AFC experiment. In the ROC paradigm, the observer will rate each lesion based on the realism of its appearance and not trivial details. Because this study was more concerned with the realistic appearance of the simulated lesions, we chose ROC methodology.

Although flexible and accurate, our simulation routines present some limitations: one concerned the lesion categories studied by the simulation routines. Although this study characterized a number of masses labeled by four relevant BI-RADS descriptors, the natural variability of masses is much larger. This same limitation was present with our microcalcification simulation routine. Another limitation concerned the

central location of lesions and the limited sized ROIs used to verify the realism of the lesion appearance. To extend the routines to a full mammogram, the lesions would need to be manually placed within the mammogram, similarly to previous work (15). Next, the mass characterization assumed a degree of elliptical symmetry in the masses. The characterization procedure would have more difficulty characterizing highly asymmetric masses. However, visual inspection of the characterized masses, even those labeled as having ill-defined borders, showed that all possessed substantial elliptical symmetry and could be characterized using our procedure. Finally, the mass model did not include the effects of mass growth. As a mass grows, it will displace the surrounding parenchymal tissue, a process that was not taken into account in the present work. Notwithstanding these limitations, however, the models produced realistic lesions as judged by experienced mammographers. Future work may address these limitations and extend the complexity of the models.

CONCLUSION

This work comprehensively measured the characteristics of common categories of benign and malignant masses and microcalcifications. The characterization measurements then directed the development of routines that could simulate the radiographic appearance of breast lesions. The simulation routines developed in this study produced masses and microcalcifications with greater complexity than existing simulation routines. In addition, observer performance experiments with experienced mammographers validated the realistic appearance of the simulated lesions.

REFERENCES

1. Pineda AR, Barrett HH. What does DQE say about lesion detectability in digital radiography? Proc SPIE 2001; 4320:561–569.
2. Barrett HH. Quality time with the fractious Fourier family [image quality]. Proc SPIE 2001; 4392:9–21.
3. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. J Opt Soc Am A 1990; 7:1266–1278.
4. Saunders RS Jr, Samei E, Hoeschen C. Impact of resolution and noise characteristics of digital radiographic detectors on the detectability of lung nodules. Med Phys 2004; 31:1603–1613.
5. Saunders JRS, Samei E, Johnson J, et al. Effect of display resolution on the detection of mammographic lesions. Proc SPIE 2005; 5749:243.
6. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353:1773–1783.
7. Chakraborty DP, Kundel HL. Anomalous nodule visibility effects in mammographic images. Proc SPIE 2001; 4324:68–76.
8. Abbey CK, Eckstein MP, Shimozaki SS, et al. Human-observer templates for detection of a simulated lesion in mammographic images. Proc SPIE 2002; 4686:25–36.
9. Huda W, Qu G, Jing Z, et al. How does observer training affect imaging performance in digital mammography? Proc SPIE 2000; 3981:259–266.
10. Johnson JP, Nafziger JS, Krupinski EA, et al. Effects of grayscale window/level parameters on breast lesion detectability. Proc SPIE 2003; 5034:462–473.
11. Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. J Opt Soc Am A 1997; 14:2420–2442.
12. Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. J Opt Soc Am A 2001; 18:473–488.
13. Carton AK, Bosmans H, Van Ongeval C, et al. Development and validation of a simulation procedure to study the visibility of micro calcifications in digital mammograms. Med Phys 2003; 30:2234–2240.
14. Krupinski EA, Johnson JP, Roehrig H, et al. Human vision model to predict observer performance: detection of microcalcifications as a function of monitor phosphor. Proc SPIE 2003; 5034:20–24.
15. Kallergi M, Gavrielides MA, He L, et al. Simulation model of mammographic calcifications based on the American College of Radiology Breast Imaging Reporting and Data System, or BIRADS. Acad Radiol 1998; 5:670–679.
16. Lefebvre F, Benali H, Gilles R, et al. A simulation model of clustered breast microcalcifications. Med Phys 1994; 21:1865–1874.
17. D'Orsi C.J. Illustrated Breast Imaging Reporting and Data System (BI-RADS). Reston, Va: American College of Radiology, 1998.
18. Heath M, Bowyer K, Kopans D, et al. The Digital Database for Screening Mammography. In: Proceedings of the 5th International Workshop on Digital Mammography. Madison, Wis: Medical Physics Publishing, 2000.
19. Marr D, Hildreth E. Theory of edge detection. Proc R Soc Lond B Biol Sci 1980; 207:187–217.
20. Samei E, Flynn MJ. An experimental comparison of detector performance for direct and indirect digital radiography systems. Med Phys 2003; 30:608–622.
21. Boone JM, Lindfors KK, Cooper VN 3rd, et al. Scatter/primary in mammography: comprehensive results. Med Phys 2000; 27:2408–2416.
22. NEMA. Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Display Standard Function. Rosslyn, Va: 2000
23. Samei E, Badano A, Chakraborty D, et al. Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. Med Phys 2005; 32:1205–1225.
24. Heiberger RM, Holland B. Statistical analysis and data display: an intermediate course with examples in S-plus, R, and SAS. New York: Springer, 2004.
25. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Stat Med 1998; 17:1033–1053.
26. Krupinski EA, Johnson J, Roehrig H, et al. Use of a human visual system model to predict the effects of display veiling glare on observer performance. Proc SPIE 2004; 5372:423–429.
27. Burgess AE, Chakraborty S. Producing lesions for hybrid mammograms: extracted tumours and simulated microcalcifications. Proc SPIE 1999; 3663:316.
28. Krupinski EA, Johnson J, Roehrig H, et al. Use of a human visual system model to predict observer performance with CRT vs LCD display of images. J Dig Imag 2004; 17:258–263.
29. Krupinski EA, Johnson J, Roehrig H, et al. On-axis and off-axis viewing of images on CRT displays and LCDs: observer performance and vision model predictions. Acad Radiol 2005; 12:957–964.
30. Yaffe MJ, Johns PC, Nishikawa RM, et al. Anthropomorphic radiologic phantoms. Radiology 1986; 158:550–552.
31. Nappi J, Dean PB, Nevalainen O, et al. Algorithmic 3D simulation of breast calcifications for digital mammography. Comput Methods Programs Biomed 2001; 66:115–124.

Resolution and noise measurements of five CRT and LCD medical displays

Robert S. Saunders, Jr.[a)]
Duke Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, North Carolina 27710, and Department of Physics, Duke University, Durham, North Carolina 27710

Ehsan Samei
Duke Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, North Carolina, 27710, and Departments of Physics and Biomedical Engineering, Duke University, Durham, North Carolina 27710

The performance of soft-copy displays plays a significant role in the overall image quality of a digital radiographic system. In this work, we discuss methods to characterize the resolution and noise of both cathode ray tube (CRT) and liquid crystal display (LCD) devices. We measured the image quality of five different commercial display devices, representing both CRT and LCD technologies, using a high-quality charge-coupled device (CCD) camera. The modulation transfer function (MTF) was calculated using the line technique, correcting for the MTF of the CCD camera and the display pixel size. The normalized noise power spectrum (NPS) was computed from two-dimensional Fourier analysis of uniform images. To separate the effects of pixel structure from interpixel luminance variations, we created structure-free images by eliminating the pixel structures of the display device. The NPS was then computed from these structure-free images to isolate interpixel luminance variations. We found that the MTF of LCDs remained close to the theoretical limit dictated by their inherent pixel size (0.85 ± 0.08 at Nyquist frequency), in contrast to the MTF for the two CRT displays, which dropped to 0.15 ± 0.08 at the Nyquist frequency. However, the NPS of LCDs showed significant peaks due to the subpixel structure, while the NPS of CRT displays exhibited a nearly flat power spectrum. After removing the pixel structure, the structured noise peaks for LCDs were eliminated and the overall noise magnitude was significantly reduced. The average total noise-to-signal ratio for CRT displays was $6.55\% \pm 0.59\%$, of which $6.03\% \pm 0.24\%$ was due to interpixel luminance variations, while LCD displays had total noise to signal ratios of $46.1\% \pm 5.1\%$ of which $1.50\% \pm 0.41\%$ were due to interpixel luminance variations. Depending on the extent of the blurring and prewhitening processes of the human visual system, the magnitude of the display noise (including pixel structure) potentially perceived by the observer was reduced to $0.43\% \pm 0.01\%$ (accounting for blurring only) and $0.40 \pm 0.01\%$ (accounting for blurring and prewhitening) for CRTs, and $1.02\% \pm 0.22\%$ (accounting for blurring only) and $0.36\% \pm 0.08\%$ (accounting for blurring and prewhitening) for LCDs. © *2006 American Association of Physicists in Medicine.* [DOI: 10.1118/1.2150777]

Key words: Image quality, Medical Display, Modulation Transfer Function, Normalized Noise Power Spectrum, Liquid Crystal Display, Cathode Ray Tube

I. INTRODUCTION

For many years, radiographic images were acquired with screen-film systems. A screen-film system bundled detection, image processing, and image display into one device. The advent of digital systems separated these functions into distinct components that could be independently optimized.[1] The image quality of a digital x-ray system, therefore, does not solely depend on the detector, but also on all components of the imaging chain, including the display device utilized.[2] In order to form a complete picture of a system's image quality, one must thoroughly measure the physical characteristics of the display device utilized.

Currently, medical displays rely on two underlying technologies. Based on an older technology, cathode ray tube (CRT) displays use a focused electron beam striking upon a phosphor to create an image.[3] In contrast, liquid crystal display (LCD) devices control the light output from individual pixels with liquid crystals and polarizing filters.[3] The resolution and noise of these display types are governed by different physical processes. The resolution of a CRT display depends on the extent and control of the electron beam. The monitor yields lower resolution at higher luminance levels and at the display peripheries, as the electron beam spreads at these luminance levels and beam projections.[4,5] Furthermore, the resolution of a CRT systematically degrades with age due to deterioration of the electron gun and a necessary increase in electron beam intensity because of a loss of phosphor luminance efficiency.[6,7] In contrast, LCDs allow for very high resolution, often approaching the limit dictated by their pixel size.[8] However, each pixel requires a significant amount of electronics to operate, which leads to considerable structured noise patterns.[9]

TABLE I. Description of the five display systems evaluated in this study. The first five rows are based on manufacturer specifications, while the next two rows reflect quantities measured in our laboratories (Ref. 25). The last row indicates the magnification ratio used for image acquisition, or the number of camera pixels used to image one display pixel.

	Barco MGD 521	Barco MGD 521M	IBM T221	National display systems Nova III	National display systems Nova V
Display Card	Barco MP1H (10-bit)	Barco 5MP2 (10-bit)	NVIDIA Quadro FX 4000 (32-bit floating point)	RealVision MD3mp (10-bit)	RealVision MD5mp (10-bit)
Type	CRT	CRT	LCD	LCD	LCD
Additional properties	p45 phosphor	p45 phosphor	Color display		
Pixel pitch (mm)	0.148	0.148	0.125	0.207	0.165
Matrix size	2048×2560	2048×2560	3840×2400	1536×2048	2048×2560
Active display area	$304\,mm \times 380\,mm$	$304\,mm \times 380\,mm$	$478\,mm \times 299\,mm$	$318\,mm \times 424\,mm$	$338\,mm \times 422\,mm$
$L_{min}(Cd/m^2)$	0.52	0.60	0.83	0.43	0.52
$L_{max}(Cd/m^2)$	308	316	235	369	371
Magnification ratio for measurement	29.6	29.6	25.0	41.4	33.0

Several researchers have considered display resolution when evaluating image quality for soft-copy displays.[7,10,11] The resolution of a display does influence the information content of an image, but other factors also affect the displayed image. For instance, investigators have more recently given attention to the noise properties of display devices.[9,12] As the magnitude and spatial frequency content of noise may impact the overall clinical utility of a display device, one must quantify both the resolution and noise of these displays to form an accurate picture of display performance.

The purpose of this work is to measure the resolution and noise properties of several medical displays, including both CRT and LCD technologies. Two key metrics were examined, the modulation transfer function (MTF) and normalized noise power spectrum (NPS), which summarize the resolution and noise properties of the display, respectively.[13–16] In addition, this paper introduces new methods for isolating the structured noise of CRTs and LCDs.

II. METHODS AND MATERIALS

A. Display description

Five different medical-grade display devices were evaluated, as listed in Table I, representing both cathode ray tube (CRT) and liquid crystal display (LCD) devices. All displays were calibrated to the Digital Imaging and Communications in Medicine (DICOM) standard according to the display manufacturer before measurements. All experiments were conducted in a room with controlled low ambient lighting set to 9 lux illuminance.

B. Camera description and evaluation

The physical characteristics of the display devices were measured using a charge-coupled device (CCD) camera (XCD-SX900, Sony Corporation, Tokyo, Japan) equipped with a macro lens (Rodgen 1:4, 28mm, Rodenstock, Munich, Germany). The camera captured images of 1280 $\times$ 960 pixels in size with a CCD chip of 6.5×4.8 mm employing a pixel size of 4.65×4.65 μm. The lens was set to

its highest magnification, such that one camera pixel imaged a $0.0050\,mm \times 0.0050\,mm$ area in the focal plane. The lens used a small aperture with a f-stop of f/11 to ensure the camera had a relatively large depth of field, which allowed objects near the true focal plane to also be captured with relative sharpness. The camera was secured on a custom gantry, offering coarse linear movement as well as fine linear movement with 0.01 mm precision (See Fig. 1). Data were transferred to a PC workstation through a FireWire connection using an image acquisition software (ImageJ; Research Services Branch, National Institute of Mental Health, Bethesda, Maryland).

To correct for any gain nonuniformities from the camera, the flat-field response of the camera was measured. As the

FIG. 1. High-quality CCD camera mounted on custom gantry for measurement of display characteristics. The gantry was capable of both coarse and fine linear movement.

gain characteristics of the camera depended on luminance, this measurement was conducted for each of the luminance levels used during display measurements. The light source consisted of a standard radiographic lightbox (X-ray Film Illuminator, S&S X-ray Products, Brooklyn, NY) covered with a neutral density filter to achieve a given luminance. Opal diffusing glass (Edmund Optics, Barrington, NJ) was placed next to the filter, which created a near Lambertian source. The camera was supplemented with a cone constructed of graphics arts black paper with velvet-type, black, light absorbing cloth. This ensured that the camera only captured light that had come through the diffuser. Finally, the diffuser was positioned several centimeters behind the camera focus; otherwise this may have revealed small nonuniformities in the diffuser, affecting the results. The camera acquired ten images at each luminance level. A gain map was formed from the average of these images. All subsequent display measurements were corrected by the appropriate gain map (corresponding to the approximate average luminance of the display) as

$$I'(x,y) = \frac{\bar{g}(L) - \beta}{G(L;x,y) - \beta}[I(x,y) - \beta], \qquad (1)$$

where $G(L;x,y)$ represents the average flat-field image at luminance L with mean $\bar{g}(L)$, $I(x,y)$ refers to the uncorrected image, β represents the pixel value at zero luminance value, and $I'(x,y)$ corresponds to the corrected image.

The inherent resolution performance of the camera was computed using the edge technique. The camera acquired an image of an edge of a 1 mm square on a glass slide resolution target (1951 USAF slide, Edmund Industrial Optics, Barrington, NJ). The slide was backlit using the same lightbox covered with a neutral density filter to achieve a luminance level of 269 cd/m^2. The MTF was calculated from the edge image using a previously published method.[17] First, a Radon transformation was applied to the data to determine the line angle with 0.01 deg accuracy. The image data were then projected along lines parallel to the edge transition, forming the edge spread function (ESF). This projection was applied in a 1.19 mm × 1.19 mm region centered on the edge and the data were placed into bins of 0.1 pixel in size. A fourth-order moving polynomial fit provided modest smoothing for the ESF while minimizing noise. The ESF was subsequently differentiated using a discrete derivative to form the line spread function (LSF). The tails of the LSF were forced to zero using a Hann window of 0.5 mm. Finally, the MTF was computed from the normalized fast Fourier transform (FFT) of the LSF.

C. Measurement of display resolution

The display resolution was measured using the line spread function (LSF) technique. The TG18-RV50 and TG18-RH50 test patterns provided vertical and horizontal line patterns, respectively.[7,18] These patterns utilized subtle lines, with 12% pixel value contrast from the background, in order to satisfy the quasilinear system requirements of the MTF measurements. The CCD camera acquired magnified images of

the displayed line pattern for each display device, where the line appeared approximately in the center of its field of view.

One caveat to the MTF measurement process concerns the concept of focus. While the camera must be in focus to capture correct information, the literature devotes few references to quantitative definitions of focus. As out of focus images are relatively blurred compared to their in-focus counterparts, the level of detail in a focused image is maximized, thus maximizing the standard deviation of the image.[11] As the camera in this study used a small aperture, it offered a relatively large depth of field. This allowed the camera to provide in focus images of LCDs that are composed of several thin, closely spaced planes of electronics and optical equipment. Experimentally, focusing was achieved by placing the camera where the image visually appeared to be in focus. The camera was then moved around that initial position sequentially until the standard deviation of the image was maximized. The image that possessed the highest standard deviation was considered to be in focus.

Our MTF measurement technique aimed to characterize the MTF of displays independent of noise properties for the display. As CRT displays and LCDs possessed different types of structured noise, the structured noise was removed from the line images using two different methods. For CRT displays, a structure map was created of the raster lines by averaging the image data along the raster line direction. The raster map was then subtracted from the line image to create a structure-free image. This procedure only averaged over areas of the image not containing the line test pattern. For vertical line patterns where the line pattern was perpendicular to the raster structure, this method could create a map of all raster lines. However, for horizontal line patterns, the line pattern was parallel to the raster lines and thus the area immediately surrounding the line pattern was excluded from this correction procedure. For LCDs, we averaged 20 pictures of the line pattern and 20 pictures of the pixel background. The average background image was then subtracted from the average line image.[11] The MTF was computed from these structure-free line images.

The MTF was calculated from the acquired line images using a modified version of the MTF calculation routine described in Sec. II B. To calculate the line angle, the image was blurred with a Gaussian kernel and then thresholded. The magnitudes for the Gaussian blur and the thresholds were determined from statistical analysis of the experimentally acquired images to give the best estimate of the line angle unaffected by noise. The angle of this thresholded line was then determined through a linear regression. Next, the pixel values of the original image were binned along lines parallel to the line pattern to form the line-spread function. This binning occurred in a 2.5 mm × 2.5 mm region centered on the line pattern with bins of one camera pixel in size. To correct for background trends in the data, a line was fit to the tails of the LSF.

Signal processing of the LSF preserved the central line area, defined as four display pixels on either side of the line peak, while processing the data in the tails of the LSF. A modified Hann window of one display pixel in width was
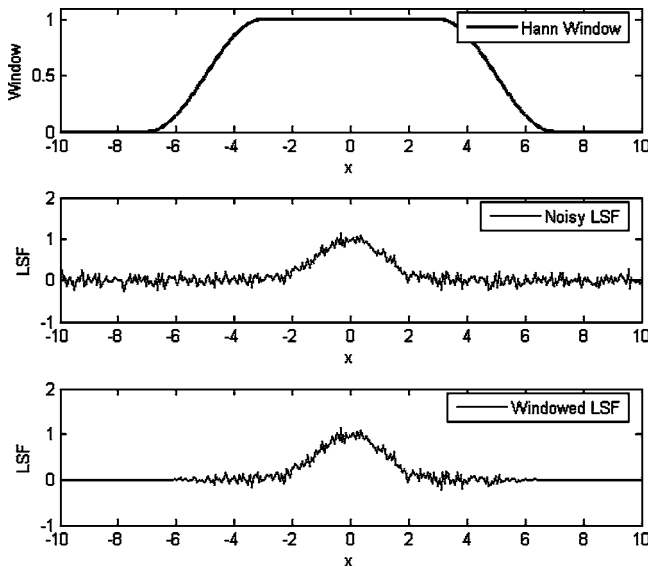
FIG. 2. Schematic of windowing procedure for the line spread function. The top curve shows a simple example of the Hann window. The middle subfigure illustrates an example noisy line spread function. The final subfigure shows the line spread function after application of the Hann window. This forces the edges of the LSF to zero to meet the criteria for Fourier analysis.

utilized to force the tails of LSF to zero, while protecting the central line area. The window took the following functional form

$$H(x) = \begin{cases} 1 & |x| < a \\ \left(\dfrac{1}{2}\right)\left(1 + \cos\left[\pi\dfrac{|x|-a}{b-a}\right]\right) & a < |x| < b \\ 0 & |x| > b \end{cases}, \quad (2)$$

where x represents the distance from the central peak of the line spread function, a denotes the length of the protected central line area (i.e., four display pixels in our routine) and $b-a$ corresponds to the distance over which the Hann window goes to zero (i.e., one display pixel in this routine). Figure 2 illustrates a simple case of applying the window function to a noisy line spread function.

Finally, the MTF was computed as the normalized FFT of the LSF. To account for the camera MTF and display pixel size, the results were divided by the MTF of the CCD camera and the sinc function corresponding to the display pixel size as

$$MTF_{display}(u) = \frac{MTF_{measured}(u)}{MTF_{camera}(u)\,\text{Sinc}(u\delta)}, \quad (3)$$

where MTF_{camera} represents the camera MTF, $MTF_{measured}$ refers to the experimentally measured MTF, δ describes the pixel size, and $MTF_{display}$ corresponds to the true MTF of the display device.

D. Measurement of display noise

The noise was evaluated using Fourier analysis of uniform images. For each display device, the camera acquired magnified images of a uniform gray area of the TG18-NS50

test pattern.[18] Similar to the resolution measurements, several preliminary images were acquired to determine whether the images were in focus. The frequency content of the image noise was evaluated in terms of the normalized noise power spectrum (NPS).[17,19] First, a region (3.8 mm $\times$ 5.1 mm) was extracted from the center of the image. This method assumed that the pixel structure in this region would be representative of the other areas of the display. This assumption should be satisfied by most displays constructed using modern manufacturing methods, producing similar pixel structures across the display. The region was then segmented into 117 overlapping regions of interest (ROIs) of 256×256 pixels (1.3 mm $\times$ 1.3 mm). The ROIs overlapped with each of their nearest neighbors by 50%. Each region was scaled by its mean pixel value to form the relative signal. A Hamming window was applied to each ROI to ensure the ROI approached zero at its edges. After computing the two-dimensional FFT of each ROI, the NPS was computed as the average of the absolute magnitude squared of each FFT.

In order to further understand the noise properties of the displays, the total noise was decomposed into two different categories following an analysis similar to a previous study.[9] This separated the total noise into two classes corresponding to different physical properties of the display: interpixel and intrapixel variations. The first category, interpixel variations, included the differences in luminance between pixels. CRT phosphor structured noise could be considered as interpixel noise, while for a LCD, such fluctuations were often caused by the nonuniform thickness of the liquid crystal elements across the display. The physical structure of the pixel caused the second form of variation, intrapixel noise. Whereas an observer would experience both forms of noise when viewing images on a display, this analysis explored how much of the total noise of a display was due to interpixel luminance variations and the pixel structure (i.e., intrapixel) components.

To isolate the interpixel luminance variations, the images were processed to remove the physical structure of the pixels or intrapixel variations. For CRT displays, the pixel structure was removed by the raster profile subtraction method (see Sec. II C). For LCDs, the following procedure was followed. The experimentally acquired uniform images were rotated to align their pixels along the horizontal direction. Due to careful camera positioning, this rotation angle remained below 1°. The rotated image was summed across both the horizontal and vertical directions. These horizontal and vertical traces showed a peak at the center of each subpixel, such that a full pixel could be constructed by counting the appropriate number of horizontal and vertical peaks. The procedure then created a pixel grid across the image, as displayed in Fig. 3(a). The routine looped through the grid and centered each grid rectangle on the pixel center. This pixel grid was visually inspected to ensure that the grid properly enclosed the pixels. An image of interpixel luminance variations was then formed where each grid rectangle was replaced by its mean luminance value. While this process removes the subpixel
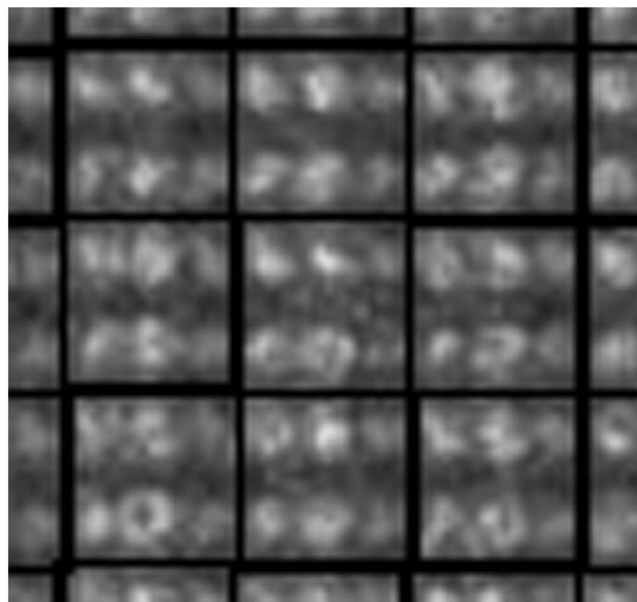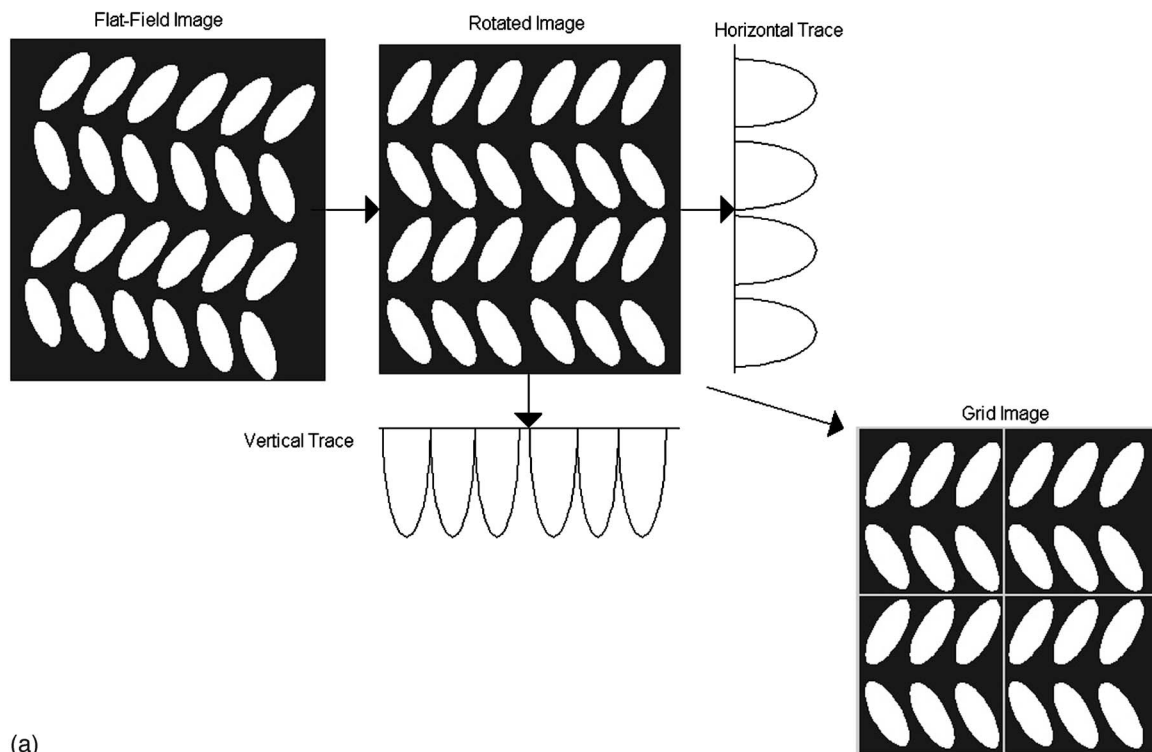
(a)

(b)

FIG. 3. Graphical description of pixel alignment procedure (a) and example of pixel alignment procedure on region of IBM T221 display (b). The dark lines indicate the borders of the pixel box.

structures for the LCD, the inherent pixelation effects associated with digital images remains. An example of this pixel alignment procedure is shown in Fig. 3(b). The NPS was recalculated from the pixel-structure-removed LCD and CRT images to examine the contribution of pixel structure to the total display noise.

III. RESULTS

Figure 4 illustrates the inherent MTF and NPS of the CCD camera. The camera provided a very high MTF over the frequency range of interest, declining only to 0.88 at 10 mm^{-1}. The MTFs of all displays were corrected by the MTF of the camera to present an accurate estimate of display resolution. However, the noise images were not corrected by the MTF, as this would unacceptably amplify the high-frequency noise.[10] The camera NPS corresponded to white noise with a very low magnitude of $5 \cdot 10^{-9}$ to 10^{-8} mm^2 over the entire frequency range of interest. This indicates that the camera added minimal noise to the acquired images.
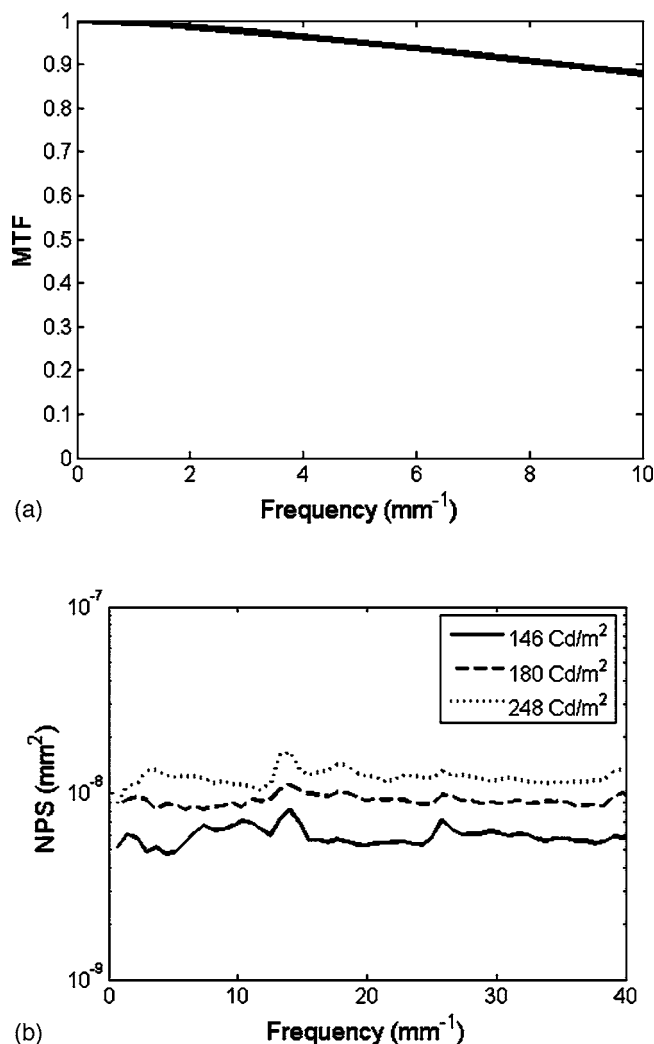
FIG. 4. Plot of the (a) MTF and (b) radial trace of the NPS of the CCD camera. The MTF remains high over the frequency range of interest. The NPS magnitude remains white and low over the entire frequency range of interest.

Figure 5 shows the measured MTF for the five display devices over the frequency range of interest from zero frequency to the Nyquist frequency dictated by the display pixel size. The first two graphs [Figs. 5(a) and 5(b)] pertain to CRT display devices while the final three plots [Figs. 5(c)–5(e)] pertain to LCDs. The LCD MTFs stayed close to unity throughout the clinically relevant frequency range of 0–4 mm^{-1}, while the MTFs for CRT displays contained far less power at higher frequencies. Each plot includes the MTF calculated along the horizontal and vertical directions in order to indicate any potential asymmetries in resolution. The horizontal and vertical MTFs remained similar for the LCDs, which indicated little asymmetry in the resolution properties of these display devices. This contrasted with the CRT displays, which exhibited notable differences between the horizontal and vertical directions, as different physical properties control the resolution in each direction.[6] As noted in Sec. II C, the horizontal MTF included some effects from the

raster line pattern, which contributed some noise to the measured MTF.

Figure 6 illustrates traces of the normalized noise power spectrum of the total system noise for five display devices. The NPS for the CRT displays showed one peak in the vertical direction corresponding to the raster line structure. In contrast, the NPS for the LCDs revealed multiple peaks from the subpixel structure. In addition, the overall noise magnitude for the CRT displays was lower than that of the LCDs. Figure 7 shows, for example, two-dimensional NPS presented in a logarithmic scale for an example CRT and an example LCD. The CRT NPS exhibited only two peaks along the vertical axis while the LCD NPS presented a complex structure across the frequency range.

Figure 8 illustrates the NPS calculated from the images after pixel structure removal. The NPS for CRTs no longer exhibited a peak in the vertical direction, as the raster structure was eliminated, while the magnitude remained largely constant. For LCDs, the overall noise magnitude dropped significantly. In addition, the shape of the NPS changed, such that the shape now resembled the sinc function corresponding to the display pixel size. Figure 9 shows two examples of two-dimensional NPS after the structure removal procedure. Compared to their counterparts in Fig. 7, these NPS of the interpixel luminance variations exhibited few peaks from the pixel structure, but peaks due to the inherent pixelation effects remained. Table II summarizes the magnitude of the noise for displays before and after the structure removal process. As expected, the pixel structure removal procedure greatly lowered the overall variance for LCDs, indicating that subpixel structure acts as the primary source of noise for LCDs. In contrast, the variance for CRT displays stayed similar to the noise variance without pixel structure removed, suggesting that interpixel luminance variations compose the primary form of noise for this display type.

IV. DISCUSSION

To fully quantify the performance of a digital x-ray imaging system, the properties of the display device must be considered. This work measured both the resolution and noise of two medical display technologies using a robust methodology for the in-field measurement of display resolution in clinical settings. The measurement procedure corrected for differing pixel structure, which isolated the structured noise from luminance variations between pixels. If implemented commercially, this methodology may be used by institutions interested in display characterization.

Our MTF calculation procedure was very similar to previous work by Samei and Flynn[7] with two notable differences. First, the line angle was computed using a linear regression of the thresholded line, as opposed to a Hough transform. The regression showed less sensitivity to the structured noise common to LCDs. Second, in order to reduce the impact of display noise on our MTF results,
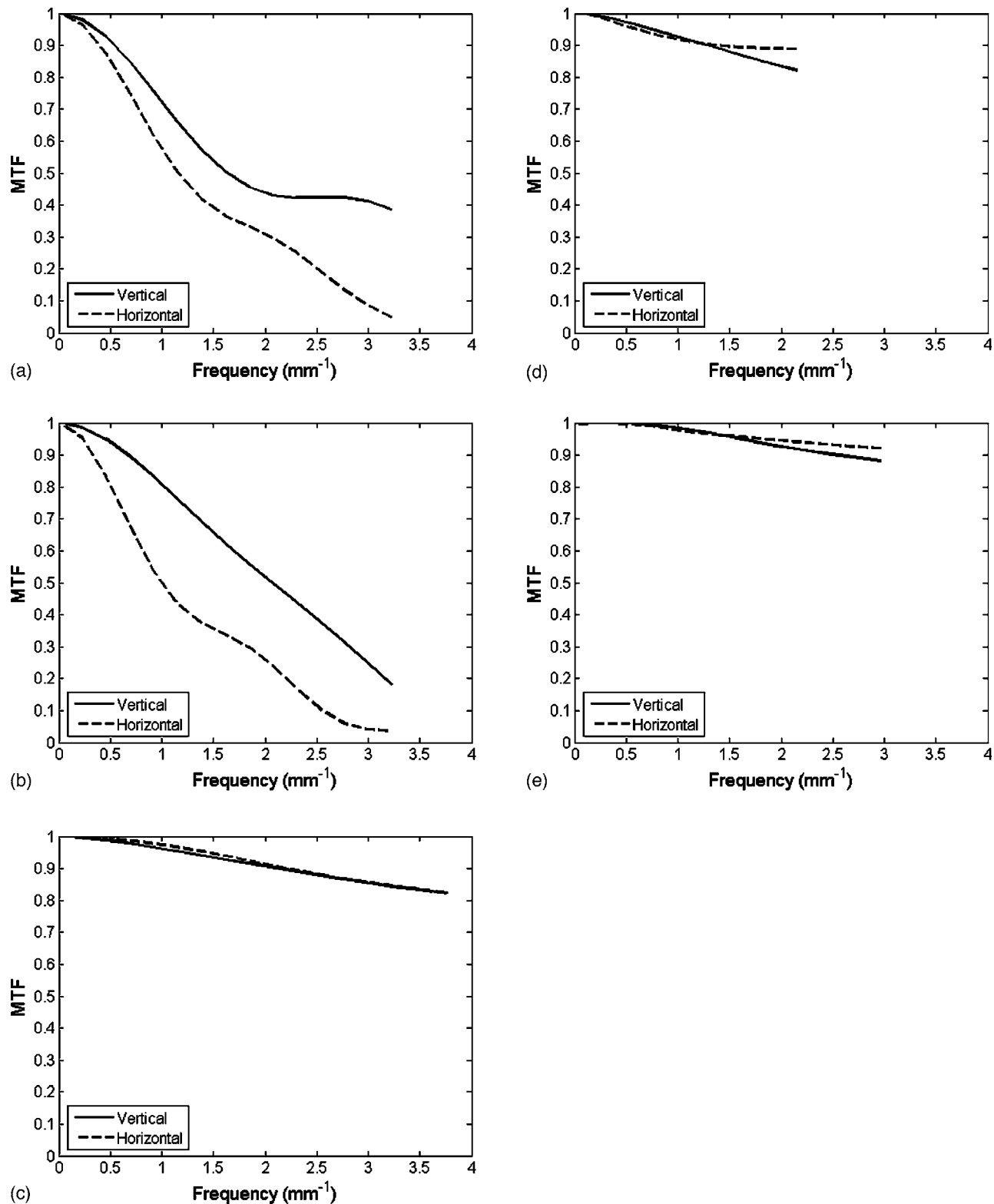
FIG. 5. Measured MTFs for (a) Barco MGD 521, (b) Barco MGD 521M, (c) IBM T221, (d) NDS Nova III, and (e) NDS Nova V displays. For the CRT displays, the horizontal and vertical MTFs diverge due to the difference in the processes impacting resolution in the two directions. For the LCDs, little asymmetry exists between the horizontal and vertical axes and the MTF remains high over the frequency range of interest.

we removed the pixel structure noise from the line pattern images. For LCDs, the structure removal was similar to that of Roehrig et al.[11] However, this methodology proved difficult to implement for CRT displays because of temporal luminance variations. This led to the use of the raster line correction procedure, which operated on a single image.

Our noise computation procedure differed from previous measurement algorithms in the following ways. Unlike
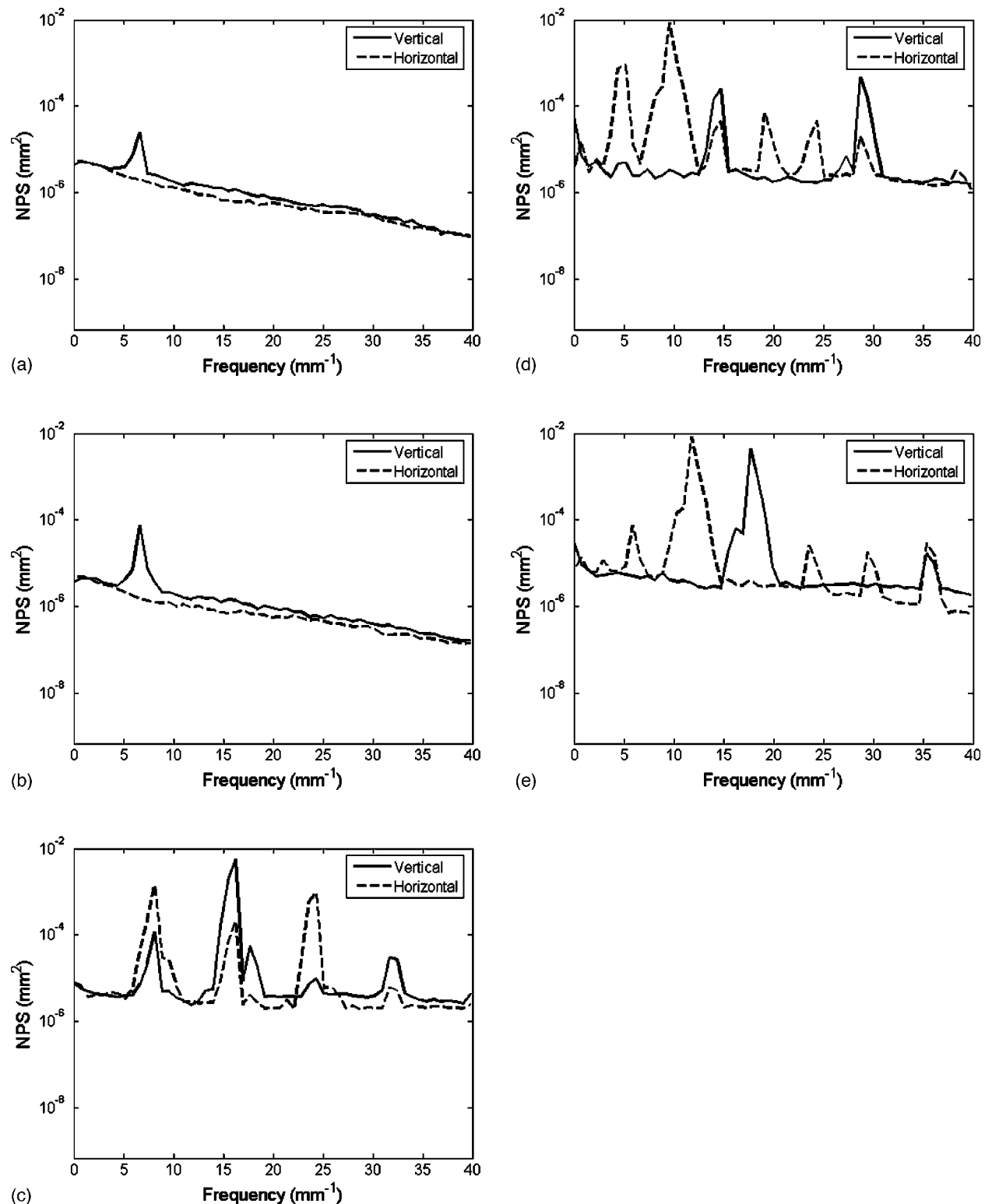
Fɪɢ. 6. Horizontal and vertical traces of the NPS of the total system noise for (a) Barco MGD 521, (b) Barco MGD 521M, (c) IBM T221, (d) NDS Nova III, and (e) NDS Nova V displays. The pixel structure causes notable peaks in the NPS for the LCD displays, while the raster structure of the CRT displays led to one peak in the vertical direction.

Muka *et al.*[20] we did not correct the uniform images by the MTF of the measurement camera, as this led to an undesirable amplification of high-frequency noise. However, we acquired all images with a narrow aperture, using only the central area of the lens and a high magnification. These two steps led to minimal resolution degradation and distortion by the lens. Similar to Badano *et al.*[9] we separated the interpixel and intrapixel noise contributions. However, we did not use
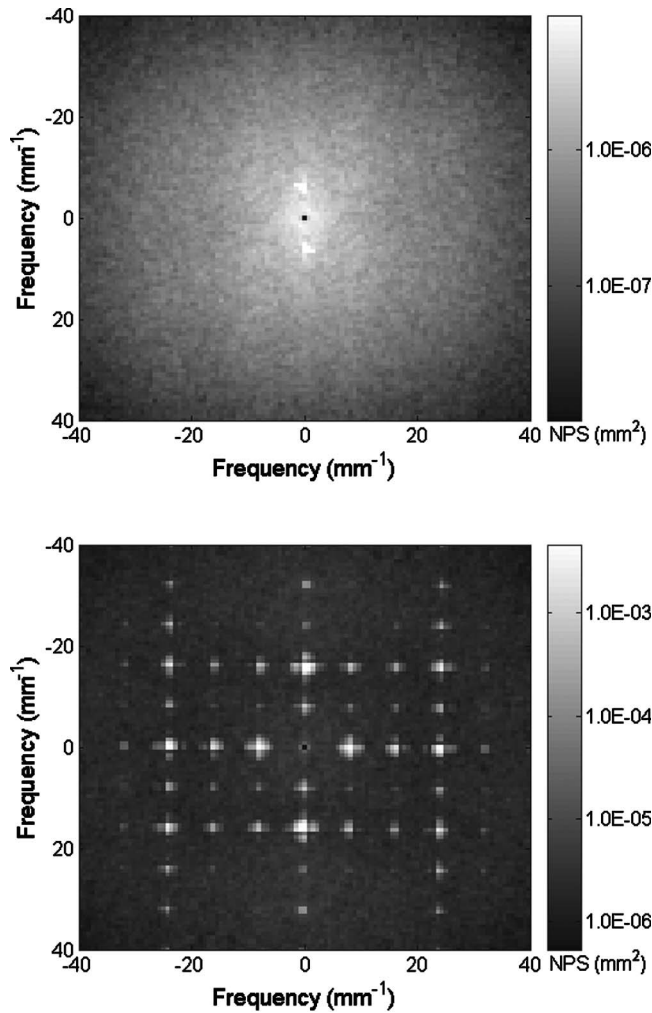
FIG. 7. Two-dimensional NPS displayed in a logarithmic scale for (a) Barco MGD 521 and (b) IBM T221 displays. The raster line leads to vertical peaks for the CRT display, while the pixel structure of the LCD produces multiple peaks across the NPS.

their pixel registration methodology to remove LCD pixel structure because of its computational cost. Instead, we examined other pixel features for LCDs to develop a pixel grid. Similar to that study, however, our pixel correction algorithm noticeably lowered the overall image noise due to the elimination of the pixel structure.

Before any measurements took place, considerable effort was devoted to characterize the properties of the CCD camera. This study included corrections for the experimentally measured MTFs by the inherent MTF of the CCD camera. At 4 mm^{-1}, the magnitude of this correction was 3.7%. In addition, careful gain calibration was performed to minimize any distortion by the lens. The magnitude of gain calibration was as high as 7.2%, with an average of 1.1%. Taken together, these two effects may have an appreciable effect on the measured MTF and NPS of a display device.

This research used a high-optical magnification to capture high-quality images of the display device. This allowed us to characterize the pixel structure with high precision, as the images showed the fine detail of the subpixel elements. In

addition, this minimized the contribution of camera blur. However, using a high-optical magnification reduced the camera field of view. Therefore, our analysis had less power in characterizing low-frequency variations often recognized as nonuniformities. This paralleled the work of previous investigators in not characterizing broad nonuniformities as noise.[9]

The NPS results showed that luminance differences between pixels constituted the primary noise source for CRT displays. The pixel structure removal eliminated the peak in the NPS, corresponding to the frequency of the raster lines, but did not alter the magnitude of the NPS. In contrast, pixel structure served as the primary noise source for LCDs. After removing structured noise, the shape of the NPS changed and the overall magnitude of the NPS dropped dramatically. This indicated that pixel structure remains the dominant source of noise for LCDs, confirming the results of Badano et al.[9] However, the pixel corrected NPS curves of CRT and LCD devices should be compared with caution as the pixel structure removal methodology differed for the two display types, due to differing pixel structures. This analysis explored what factor, interpixel luminance variations or pixel structure (intrapixel variations), represented the primary source of noise for each display type.

To understand the magnitude of the noise levels in Table II, these metrics can be compared to the quantum noise level in clinical images. For instance, the noise levels in representative mammograms and chest radiographs, including quantum noise and electronic noise, is approximately 1%–3% in terms of the standard deviation to the mean image grayscale value. The display noise values, as summarized in Table II, are comparable to these figures. This illustrates the importance and potential impact of display noise on diagnostic performance.

The noise-to-signal ratios calculated in Table II contain all noise in the image. However, two processes could reduce the impact of noise on human perception. First, there have been indications that human observers can prewhiten structured patterns from images, thus reducing their potential impact.[21] In the case of total prewhitening, the right columns of Table II would be more representative of display noise than the left columns. However, it is uncertain to what extent humans can prewhiten the structured noise of display devices. Second, human observers do not perceive the different spatial frequencies of a scene with equal acuity. One can estimate how much of the display noise could be perceived by a human observer by filtering the measured NPS results with the human visual response function $V(\rho)$ (Ref. 22) as

$$\text{NPS}_{\text{filtered}}(u,v) = \text{NPS}_{\text{measured}}(u,v)|V(\rho)|^2,$$

$$V(\rho) = |\eta \rho^{a_1} \cdot e^{-a_2 \rho^{a_3}}|^2 \qquad (4)$$

where ρ describes the radial spatial frequencies of the image in cycles/millimeters assuming a viewing distance of 40 cm, η normalizes $V(\rho)$ to one as its maximum value, and parameters (a_1, a_2, a_3) equal (1.5, 3.22, 0.68). The areas under the filtered NPS can be used as a measure of perceived noise.
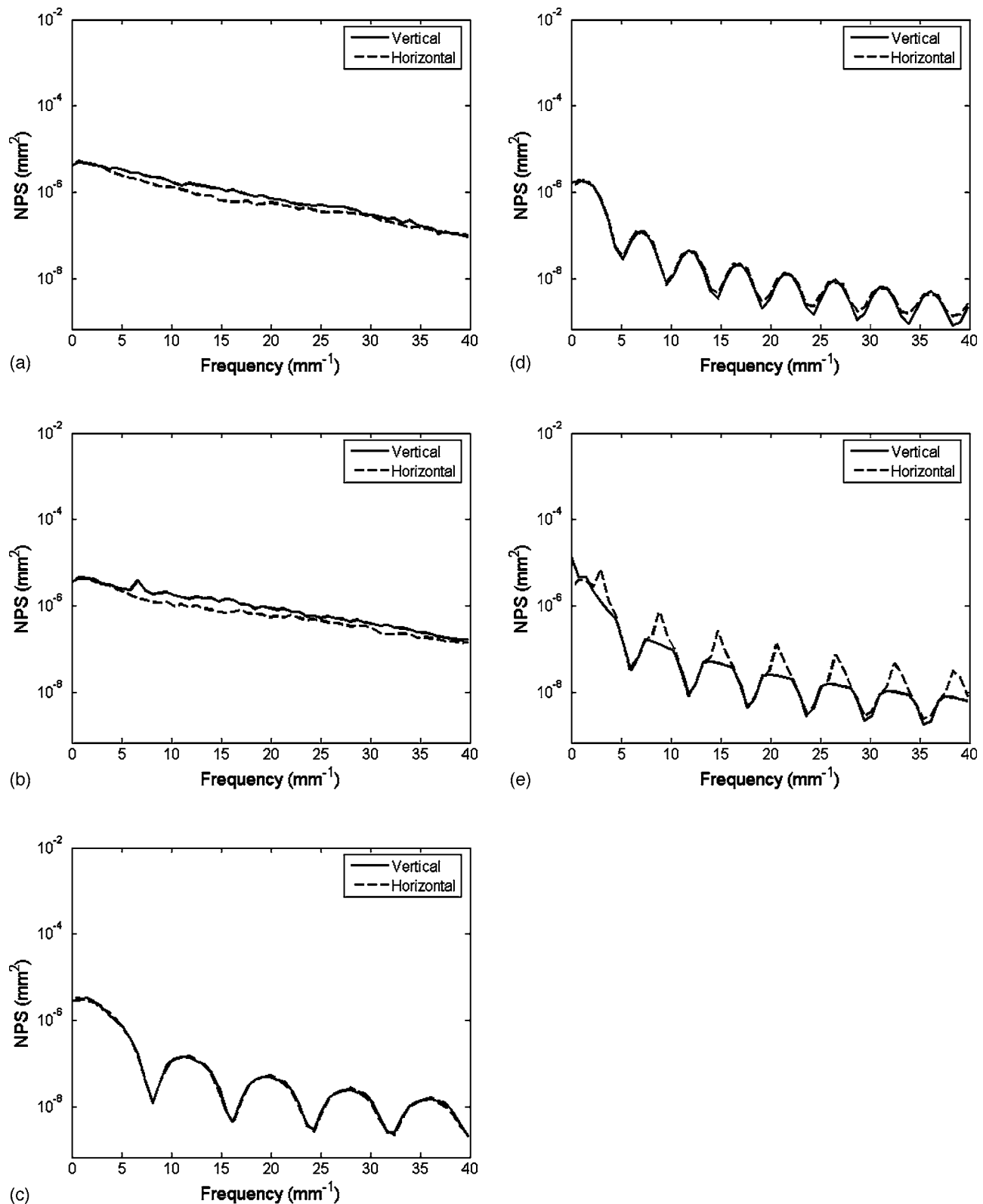
FIG. 8. After correcting for pixel structure, the noise variance drops dramatically. This may be seen in the horizontal and vertical traces of the NPS for (a) Barco MGD 521, (b) Barco MGD 521M, (c) IBM T221, (d) NDS Nova III, and (e) NDS Nova V displays.

The results, shown in Table III, indicate that the majority of pixel structured noise of LCDs will be blurred by the human visual system. The blurring was more effective for the nine megapixel LCD tested given its smaller pixel structure.

Considering the extent of possible prewhitening and frequency filtering processes, the above analysis only serves as a preliminary step in understanding the visual relevance of display noise. As the quantum noise figures noted previously
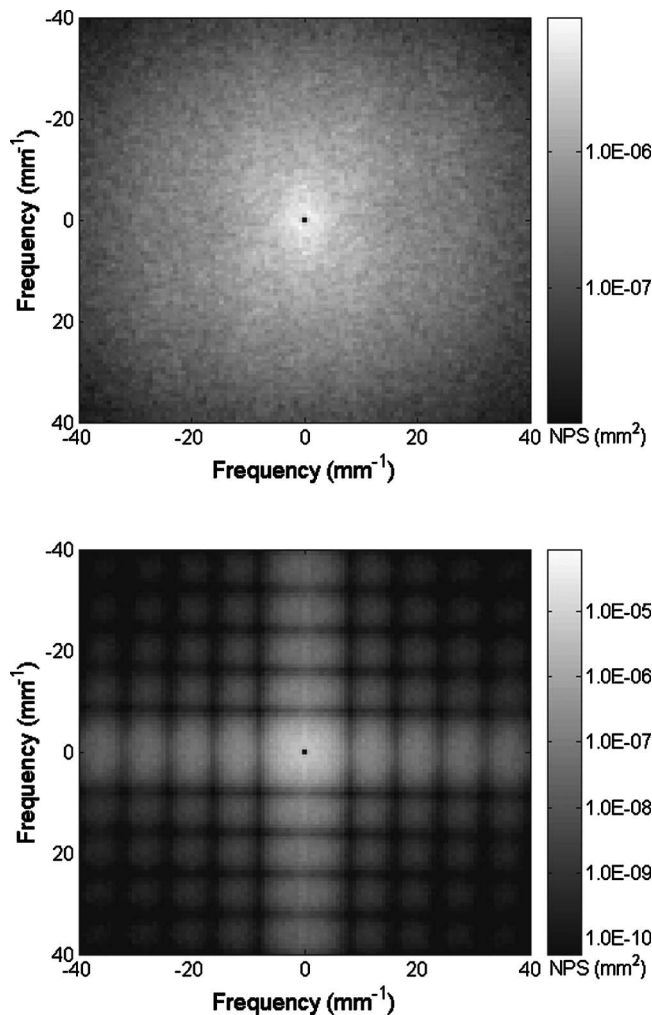
FIG. 9. Two-dimensional NPS calculated from the interpixel luminance noise displayed in a logarithmic scale for (a) Barco MGD 521 and (b) IBM T221 displays. The structure removal procedure eliminates many of the NPS peaks. While this removed the subpixel structure for the LCD display, the inherent pixelation effects remain, as evidenced by the low amplitude regular peaks.

do not compensate for the human visual response, these new noise-to-signal ratios for displays cannot be directly compared to detector noise levels. In addition, these noise figures are not reduced to detectability indices for specific clinical

TABLE II. Noise-to-signal ratio (standard deviation divided by the mean) for the CRT and LCD displays before and after the pixel structure removal procedure. These numbers were computed from the two-dimensional NPS using Parseval's theorem.

Manufacturer and model	Noise-to-Signal Ratio ($\sigma/\langle x \rangle$)	
	Without pixel structure removal (%)	With pixel structure removal (%)
Barco MGD 521	6.13	5.86
Barco MGD 521M	6.97	6.20
IBM T221	42.45	1.67
NDS Nova III	43.81	1.03
NDS Nova V	51.88	1.80

TABLE III. Perceivable noise-to-signal ratio. After compensating for the transfer properties of the human eye, the noise-to-signal ratios decrease dramatically. This suggests that the human observer may not perceive much of the structured noise from the display devices. These numbers were computed from the two-dimensional NPS filtered by the human visual response and using Parseval's theorem.

Manufacturer and model	Noise-to-Signal Ratio ($\sigma/\langle x \rangle$)	
	Without pixel structure removal (%)	With pixel structure removal (%)
Barco MGD 521	0.44	0.41
Barco MGD 521M	0.42	0.39
IBM T221	0.78	0.36
NDS Nova III	1.20	0.29
NDS Nova V	1.08	0.44

tasks.[23,24] Nonetheless, physical measurements, as undertaken in this study, form a necessary first step in characterizing a display system. Our future work will include observer experiments in order to more fully understand how the resolution and noise characteristics of displays affect clinical performance.[9,24]

V. CONCLUSIONS

This paper reports an assessment of image quality for five different commercial display devices representing both CRT and LCD technologies. The findings confirm that LCDs offer higher MTFs than CRT displays. Yet, the resolution advantages of LCDs must be considered in light of their noise properties. The CRT displays show a lower MTF, but also demonstrate lower noise. Finally, this study introduces a new means of isolating interpixel variations for both CRT and LCD devices, which will facilitate the noise comparison between monitors using different pixel structures.

ACKNOWLEDGMENTS

a)Electronic mail: Saunders@phy.duke.edu
[1] R. S. Saunders, J., E. Samei, J. L. Jesneck, and J. Y. Lo, "Physical characterization of a prototype selenium-based full field digital mammography detector," Med. Phys. **32**, 588–599 (2005).
[2] J. Fan, W. J. Dallas, H. Roehrig, and E. A. Krupinski, "Improving visualization of digital mammograms on the CRT display system," Proc. SPIE **5029**, 746–753 (2003).
[3] H. Roehrig, J. Fan, A. Chawla, and K. Gandhi, "The liquid crystal display (LCD) for medical imaging in comparison with the cathode ray tube display (CRT)," Proc. SPIE **4786**, 114–131 (2002).
[4] T. Mertelmeier "Why and how is soft copy reading possible in clinical practice?," J. Digit Imaging **12**, 3–11 (1999).
[5] E. Muka, H. R. Blume, and S. J. Daly, "Display of medical images on CRT soft-copy displays: A tutorial," Proc. SPIE **2431**, 341–359 (1995).

[6] K. D. Compton, "Factors affecting CRT display performance: Specifying what works," Proc. SPIE **3976**, 412–423 (2000).

[7] E. Samei and M. J. Flynn, "A method for in-field evaluation of the modulation transfer function of electronic display devices," Proc. SPIE **4319**, 599–607 (2001).

[8] H. R. Blume, P. M. Steven, M. E. Cobb, A. M. K. Ho, F. Stevens, S. Muller, H. Roehrig, and J. Fan, "Characterization of high-resolution liquid crystal displays for medical images," Proc. SPIE **4681**, 271–292 (2002).

[9] A. Badano, R. M. Gagne, R. J. Jennings, S. E. Drilling, B. R. Imhoff, and E. Muka, "Noise in flat-panel displays with subpixel structure," Med. Phys. **31**, 715–723 (2004).

[10] A. Badano, S. J. Hipper, and R. J. Jennings, "Luminance effects on display resolution and noise," Proc. SPIE **4681**, 305–313 (2002).

[11] H. Roehrig, J. Gaskill, J. Fan, A. Poolla, and C. Martin, "In-field evaluation of the modulation transfer function of electronic display devices," Proc. SPIE **5367**, 456–463 (2004).

[12] J. Fan, W. J. Dallas, H. Roehrig, E. A. Krupinski, K. Gandhi, and M. K. Sundareshan, "Spatial noise of high-resolution liquid-crystal displays for medical imaging: Quantitative analysis, estimation, and compensation," Proc. SPIE **5367**, 433–443 (2004).

[13] H. Fujita, K. Doi, and M. L. Giger, "Investigation of basic imaging properties in digital radiography.6. Mtfs of II-TV digital imaging-systems," Med. Phys. **12**, 713–720 (1985).

[14] C. D. Bradford, W. W. Peppler, and J. M. Waidelich, "Use of a slit camera for MTF measurements," Med. Phys. **26**, 2286–2294 (1999).

[15] M. B. Williams, P. A. Mangiafico, and P. U. Simoni, "Noise power spectra of images from digital mammography detectors," Med. Phys. **26**, 1279–1293 (1999).

[16] A. Workman and D. S. Brettle, "Physical performance measures of radiographic imaging systems," Dentomaxillofac Radiol., **26**, 139–146 (1997).

[17] R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med. Phys. **30**, 3006–3017 (2003).

[18] E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D. M. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R. A. Uzenoff, J. Wang, and C. E. Willis, "Assessment of display performance for medical imaging systems: Executive summary of AAPM TG18 report," Med. Phys. **32**, 1205–1225 (2005).

[19] M. J. Flynn and E. Samei, "Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems," Med. Phys. **26**, 1612–1623 (1999).

[20] E. Muka, T. Mertelmeier, R. M. Slone, and E. Senol, "Impact of phosphor luminance noise on the specification of high-resolution CRT displays for medical imaging," Proc. SPIE **3031**, 210–221 (1997).

[21] E. Samei, "AAPM/RSNA Physics tutorial for residents: Technological and psychophysical considerations for digital mammographic displays," Radiographics **25**, 491–501 (2005).

[22] M. P. Eckstein, J. L. Bartroff, C. K. Abbey, J. S. Whiting, and F. O. Bochud, "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," Opt. Express **11**, 460–475 (2003).

[23] E. Samei, M. J. Flynn, and W. R. Eyler, "Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs," Radiology **213**, 727–734 (1999).

[24] R. S. Saunders, Jr., E. Samei, and C. Hoeschen, "Impact of resolution and noise characteristics of digital radiographic detectors on the detectability of lung nodules," Med. Phys. **31**, 1603–1613 (2004).

[25] E. Samei and S. L. Wright, "Luminance and contrast performance of liquid crystal displays for mammographic applications," Technol. Cancer Res. Treat. **3**, 429–436 (2004).

A Monte Carlo Investigation on the Impact of
Scattered Radiation on Mammographic Resolution and Noise

Robert S Saunders, Jr [a,b], Ehsan Samei [a,b,c]
[a] Duke Advanced Imaging Labs, Department of Radiology,
Duke University Medical Center
[b] Department of Physics, Duke University
[c] Department of Biomedical Engineering, Duke University

ABSTRACT

Scattered radiation plays a significant role in mammographic imaging, with scatter fractions over 50% for larger, denser breasts. For screen-film systems, scatter primarily affects the image contrast, reducing the conspicuity of subtle lesions. While digital systems can overcome contrast degradation, they remain susceptible to scatter's impact on the image resolution and noise. To better understand this impact, we have created a Monte Carlo model of a mammographic imaging system adaptable for different imaging situations. This model flags primary and scatter photons and therefore can produce primary-only, scatter-only, or primary plus scatter images. Resolution was assessed using the edge technique to compute the Modulation Transfer Function (MTF). The MTF of a selenium detector imaged with a 28 kVp Mo/Mo beam filtered through a 6 cm heterogeneous breast was 0.81, 0.0002, and 0.65 at 5 mm^{-1} for the primary beam, scatter-only, and primary plus scatter beam, respectively. Noise was measured from flat-field images via the noise power spectrum (NNPS). The NNPS-exposure product using the same imaging conditions was $1.5 \cdot 10^{-5}$ $mm^2 \cdot mR$, $1.6 \cdot 10^{-5}$ $mm^2 \cdot mR$, and $1.9 \cdot 10^{-5}$ $mm^2 \cdot mR$ at 5 mm^{-1} for the primary, scatter, and primary plus scatter beam, respectively. The results show that scatter led to a notable low-frequency drop in the MTF and an increased magnitude of the NNPS-exposure product. (This work was supported in part by USAMRMC W81XWH-04-1-0323.)

Keywords: Image Quality, Mammography, Simulation, Monte Carlo, Modulation Transfer Function, MTF, Noise Power Spectrum, NNPS

1. INTRODUCTION

Scattered radiation has a significant impact on image quality in medical imaging. For mammographic imaging, previous studies have estimated that 50% of all photons reaching the detector when imaging large, dense breasts are scattered photons.[1] Scatter's effects depend on the particular x-ray detector used. For screen-film detectors, scatter diminishes the conspicuity of subtle lesions by reducing the image contrast. These contrast limitations are not faced by digital mammography. Digital mammography is affected, however, by scatter's effects on image resolution and noise. To measure the magnitude of these effects, this study examines system resolution and noise with and without the presence of scatter in a variety of imaging situations. By computing scatter properties, mammography detectors can be designed to more effectively reduce the deleterious effects of scattered radiation.

2. METHODS AND MATERIALS

2.1 Monte Carlo Description

To isolate the effects of scatter and primary radiation, this study used simulation methods. It simulated the photon transport physics using Penelope Monte Carlo code (version 2005).[2] Penelope performs accurate simulation of the physical photon interactions through use of both numerical databases and analytical cross-sections. Penelope has been proven accurate for electrons, positrons, and photons in the range of 50 eV to 1 GeV.[3]

The Monte Carlo was used to form a model of a direct flat-panel mammography system. This model, as shown in Figure 1, consisted of an anode, breast phantom, and a selenium detector. For resolution studies, a tungsten edge was positioned on top of the breast in order to compute an edge spread function. In addition, for some runs an antiscatter grid was located on top of the detector to explore the effects of these devices. To ensure the realism of this model, published data was used to set the physical properties for the photons, material compositions, and attenuation. The photons were emitted from the anode according to an angular distribution based on previous work.[4] The photon energies were distributed according to previously measured bremsstrahlung distributions filtered by the tube filtration.[5-7] The molecular composition of glandular material was provided by previous publications,[8] while the composition of adipose tissue was provided by Penelope.[9] Attenuation data for all materials was provided by Penelope.
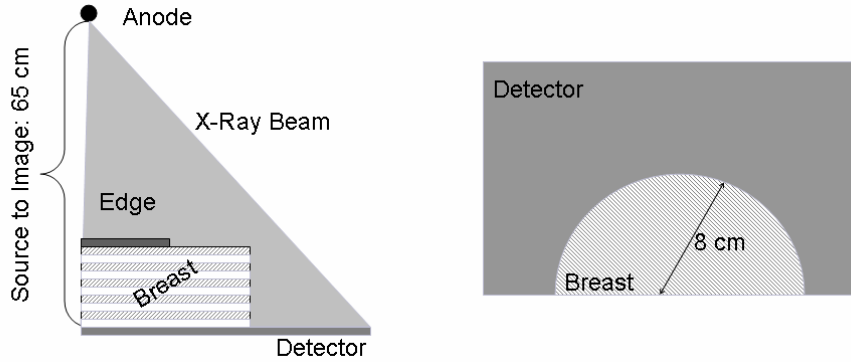

Figure 1. Schematic of simulated imaging system. In this case, the breast has a heterogeneous composition, such that the breast is composed of ten interleaving slabs of glandular and adipose tissue. The tungsten edge is used for assessing resolution, but is removed for noise evaluation.

Once a photon underwent a scattering event, such as coherent scatter or incoherent/Compton scatter, the photon was labeled as a scattered photon. Any secondary particles created from an interaction also were labeled as scattered photons. By using this labeling, the code could produce images containing only primary photons, only scattered photons, or both primary and scattered photons. If a photon interacted with the detector, the code would track the electrons produced and record the electron's position and energy. The positions were binned into pixels of 0.05 mm and the energy was integrated to produce the image signal. In addition, the code recorded the energy spectrum of all photons impinging upon the detector, regardless of whether these photons were recorded by the detector.

To efficiently investigate the effects of different model parameters, we established a default case, as shown in Table I. The effect of a specific parameter was investigated by setting all other parameters to their default value and varying only that one parameter. For instance, to explore the effects of different beam energies, all other parameters were held constant (breast composition, anode type, breast thickness, breast location, and grid status) and only the energy of the x-ray beam was varied.

Table I. Range of Simulation Parameters. The effects of specific parameters are investigated by using default values for all other parameters and varying that specific parameter.

Parameter	Default Value	Range of Values
Breast Composition	Heterogeneous	100% Adipose, Heterogeneous, 100% Glandular
Grid Status	No Grid	No Grid, Mammographic Grid
Beam Energy	28 kVp	25 kVp, 28 kVp, 32 kVp, 35 kVp
Location	Breast Center	Chest Wall, Breast Center, Nipple
Breast Thickness	6 cm	2 cm, 4 cm, 6 cm, 8 cm
Tube	Mo/Mo	Mo/Mo, W/Rh

To further model mammographic systems, all images were gain corrected to account for intensity variations. Emulating commercial systems, 10 images were acquired of a 4 cm Lucite block placed at the tube side of the system (63 cm from the detector). The 10 images were averaged together to form the gain map. All images were corrected by the appropriate gain map as:

$$I'(x,y) = \frac{\bar{G}}{G(x,y)} \cdot I(x,y) \qquad (1)$$

where I represents the input image, I' corresponds to the corrected image, and G is the average of the 10 gain images with mean $\bar{G}$. There was no offset correction as the simulated system had zero offset: an image acquired at zero exposure would have zero signal everywhere.

2.2 Resolution and Noise Assessment

Resolution was assessed through the Modulation Transfer Function (MTF).[10] This was accomplished using modified versions of established assessment routines.[11-13] Briefly, the routine went through the following steps. The routine first smoothed the image with a Gaussian smoothing kernel to reduce noise and then used a Sobel method to find the edge transition. The edge angle and intercept were determined through a linear regression. However, as the edge angle was known *a priori* for these simulation studies, that parameter was entered in manually. By binning the data along lines parallel to the edge transition, the edge spread function (ESF) was computed. As opposed to previous publications, in this work the line spread function was not computed using a finite difference, as this was overly sensitive to noise. Rather, the LSF was found from a third order moving polynomial fit. After computing the polynomial fit for an area around a given point, the derivative of that fit became the value of the line spread function for that point. Figure 2 shows examples of this polynomial differentiation compared to finite difference techniques.

Figure 2. Examples of different differentiation methods without noise (left) and with modest noise (right). The top plot shows an edge spread function with its associated line spread functions underneath. Without noise, finite difference methods and the polynomial method gave similar answers (differing by 0.3% over the range from -10 mm to 10 mm). However, in the presence of moderate noise, the two methods gave dramatically different answers. The polynomial method produced a similar LSF to the case without noise, while the finite difference method produced a substantially noisier LSF in which the line peak is barely visible.

Next, the resolution assessment routine smoothed the tails of the LSF to lower the noise of the MTF while preserving the central area of the line spread function. This preserved the shape of the MTF, as the MTF shape is determined by the width of the line spread function peak, but the smoothing removed significant amounts of noise. Finally, the LSF was transformed by a Fast Fourier Transform (FFT), normalized by its value at zero frequency, and the MTF was computed as the absolute value of that quantity.

Noise was measured by the Noise Power Spectrum (NNPS).[14-16] The images were segmented into 49 overlapping ROIs that measured 6.4 mm x 6.4 mm in size. The routine subtracted off the mean of each ROI and then normalized each by their mean and the pixel size. Each ROI was scaled by the ratio of its mean to the mean of the ROI in the top-left hand corner, to minimize the influence of intensity variations across the image. Each ROI was transformed by an FFT, averaged together, and normalized to form the NNPS. Profiles of the NNPS were taken in the radial, horizontal, vertical, and axial directions by averaging a ±5 pixel wide band through the NNPS. The NNPS were then multiplied by their exposure, as the NNPS of a linear system should scale linearly with exposure. This would discriminate between situations where the NNPS is low because it was acquired at a lower dose or because the imaging parameters used led to lower noise.

3. RESULTS

Figure 2 shows the energy spectrum of the photons reaching the detector for the default simulation case (6 cm heterogeneous breast, 28 kVp, Mo/Mo tube, no grid, center of breast). Figure 3 also shows the energy spectrum of the photons reaching the detector, including both primary and scatter, for varying beam energies. For each beam energy, the photon energy spectrum appears roughly similar for photons below 20 keV, with higher energy beams showing more photons at higher energies. Table II illustrates the scatter fractions for various beam energies. Similar to previous work,[1] scatter fractions appeared roughly constant with increasing energy.
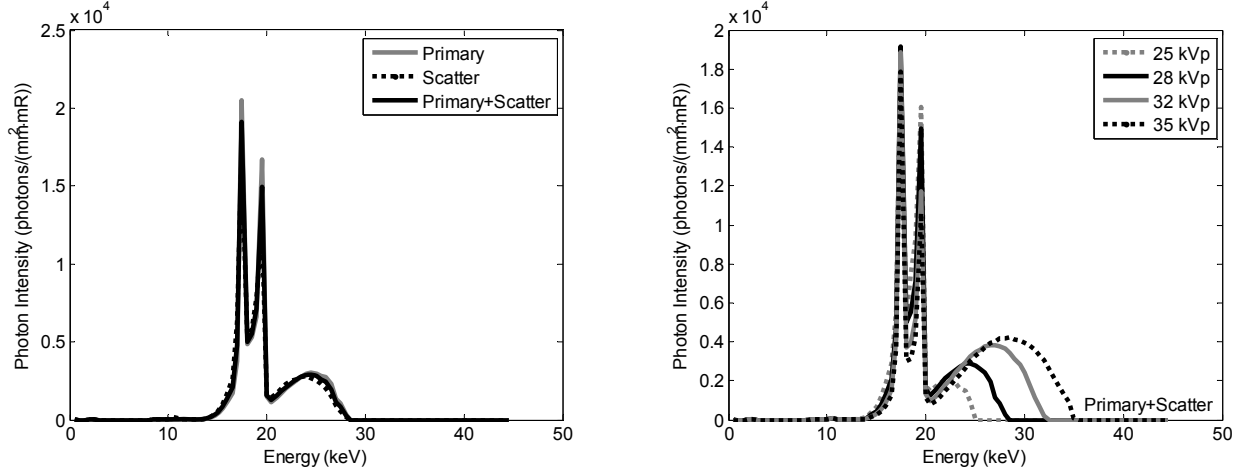

Figure 3. Normalized energy spectrum of photons, with primary-only, scatter-only, and primary plus scatter cases, reaching the detector for the default simulation case (left). Energy spectrum of all photons reaching the detector (primary plus scatter) for varying beam energies, keeping all other parameters constant (right).

Table II. Scatter fraction for various beam energies. The scatter fraction stays roughly constant with beam energy.

Beam Energy (kVp)	Scatter Fraction
25	0.387
28	0.387
32	0.386
35	0.385

Figure 4 illustrates the resolution and noise for the default simulation case. Scattered photons caused a low-frequency drop in the MTF, but also slightly changed the shape of the MTF at higher frequencies. The scattered photons act like a large blurring kernel, as indicated by its very low MTF. For the noise, scattered photons decreased the signal to noise ratio of the images, as NNPS multiplied by exposure increased between the primary-only case and the primary plus scatter case. Figure 5 shows the resolution and noise for different beam energies. The MTF and NNPS appear roughly constant across beam energies.
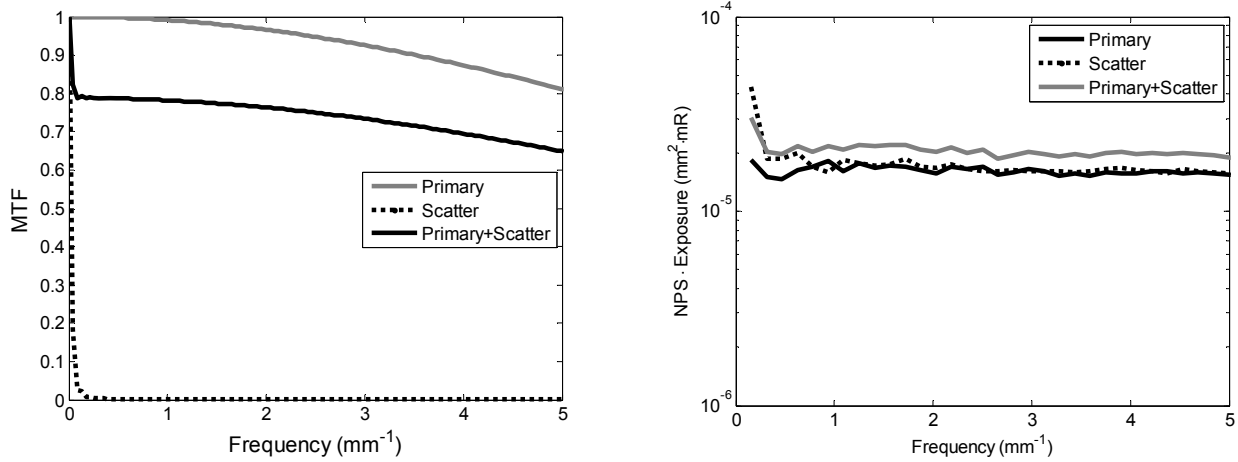
Figure 4. Resolution (left) and noise (right) for the default simulation case for primary photons only, scattered photons only, and primary plus scattered photons. Noise is represented by the radial trace of the NNPS multiplied by exposure, as the NNPS of a linear system should be inversely proportional to exposure.
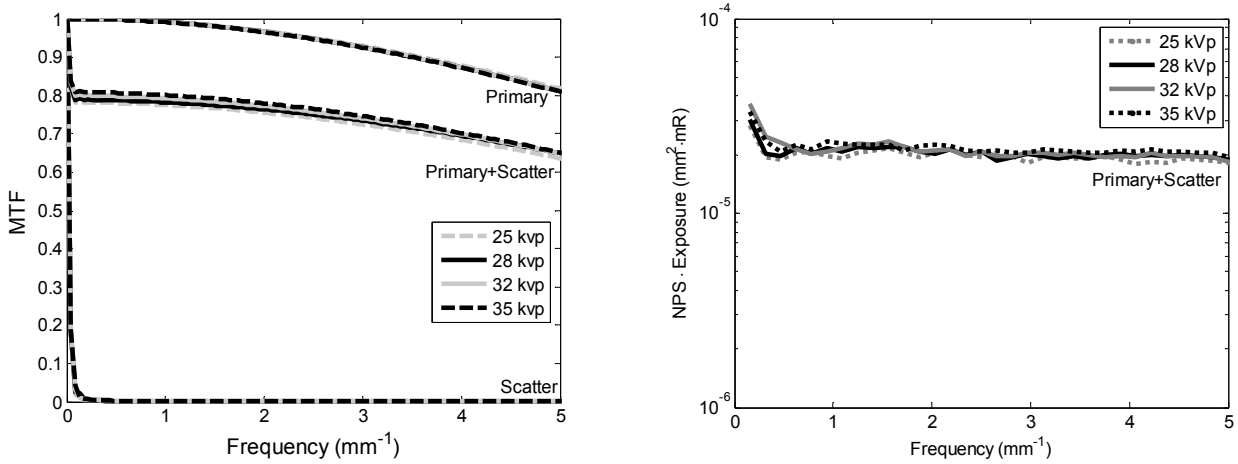

Figure 5. Resolution (left) and noise (right) for different beam energies, while controlling all other simulation parameters. The MTFs are plotted for the primary, scatter, and primary plus scatter cases, while the noise metric, the radial trace of the NNPS multiplied by exposure, only represent the noise for the primary plus scatter cases.

4. DISCUSSION AND CONCLUSIONS

Several previous investigations have modeled the scatter in radiographic systems, but have focused only on scatter fractions, contrast improvement, or signal to noise ratios.[1,17-23] A limited number of investigations have examined some aspect of the resolution and noise effects of scatter.[24-26] However, no previous work has comprehensively examined the resolution and noise effects of scattered radiation.

This study examined the resolution and noise of an imaging system both with and without the presence of the scatter. The results show how scatter affects the frequency content of images. For the MTF, scatter leads to a low-frequency drop but also changes the shape of the MTF, especially at higher frequencies. For noise, scattered photons add considerable noise to the image, leading to NNPS-exposure products with greater magnitudes.

Several items are planned for future work. The first step would be to record the glandular dose for each imaging situation. Glandular dose would allow researchers to weight the resolution and noise advantages versus the dose given to the patient. Second, the model will incorporate more scatter rejection devices, especially slot-scan devices, to expand the model utility. Finally, these results should be compared against measured results to ensure the validity of the model.

ACKNOWLEDGEMENTS

REFERENCES

[1]   J. M. Boone, K. K. Lindfors, V. N. Cooper, 3rd et al.,"Scatter/primary in mammography: comprehensive results," Med Phys **27**, 2408-2416 (2000).

[2]   F. Salvat, J.-M. Fernandez-Varea, E. Acosta et al., *PENELOPE—A Code System for Monte Carlo Simulation of Electron and Photon Transport*. (OECD Publications, Paris, France, 2001).

[3]   J. Sempau, J. M. Fernandez-Varea, E. Acosta et al.,"Experimental benchmarks of the Monte Carlo code PENELOPE," Nucl Instrum Methods B **207**, 107-123 (2003).

[4]   M. R. Ay, M. Shahriari, S. Sarkar et al.,"Monte carlo simulation of x-ray spectra in diagnostic radiology and mammography using MCNP4C," Phys Med Biol **49**, 4897-4917 (2004).

[5]   J. M. Boone, T. R. Fewell, and R. J. Jennings,"Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography," Med Phys **24**, 1863-1874 (1997).

[6]   J. Hubbell and S. M. Seltzer (2004), *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients* [Online] Available: http://physics.nist.gov/xaamdi [January 20, 2006]. National Institute of Standards and Technology, Gaithersburg, MD

[7]   J. Schweppe, X-Ray Attenuation and Energy Absorption (Wolfram Information Center, Champaign, IL, 2002).

[8]   International Commission on Radiation Units and Measurements., *Tissue substitutes in radiation dosimetry and measurement*. (International Commission on Radiation Units and Measurements, Bethesda, Md., U.S.A., 1989).

[9]   M. J. Berger, ESTAR, PSTAR, and ASTAR: Computer Programs for Calculating Stopping-Power and Ranges for Electrons, Protons, and Helium Ions, NIST Report NISTIR-4999,Washington, DC 1992.

[10]   K. Rossmann,"Point spread-function, line spread-function, and modulation transfer function. Tools for the study of imaging systems," Radiology **93**, 257-272 (1969).

[11]   R. S. Saunders and E. Samei,"A method for modifying the image quality parameters of digital radiographic images," Med Phys **30**, 3006-3017 (2003).

[12]   R. S. Saunders, Jr., E. Samei, J. L. Jesneck et al.,"Physical characterization of a prototype selenium-based full field digital mammography detector," Med Phys **32**, 588-599 (2005).

[13]   R. S. Saunders and E. Samei,"Resolution and noise measurements of five CRT and LCD medical displays," Med Phys **33**, 308-319 (2006).

[14]   M. L. Giger, K. Doi, and H. Fujita,"Analysis of Noise Wiener Spectra in Digital Ii/Tv Imaging-Systems," Med Phys **11**, 385-385 (1984).

[15]   M. L. Giger, K. Doi, and C. E. Metz,"Investigation of Basic Imaging Properties in Digital Radiography.2. Noise Wiener Spectrum," Med Phys **11**, 797-805 (1984).

[16]   M. L. Giger, K. Doi, and H. Fujita,"Investigation of Basic Imaging Properties in Digital Radiography.7. Noise Wiener Spectra of II-TV Digital Imaging-Systems," Med Phys **13**, 131-138 (1986).

[17]   Z. Jing, W. Huda, and J. K. Walker,"Scattered radiation in scanning slot mammography," Med Phys **25**, 1111-1117 (1998).

[18]   H. P. Chan, K. L. Lam, and Y. Z. Wu,"Studies of performance of antiscatter grids in digital radiography: effect on signal-to-noise ratio," Med Phys **17**, 655-664 (1990).

[19]   P. J. Papin and P. S. Rielly,"Monte Carlo simulation of diagnostic x-ray scatter," Med Phys **15**, 909-914 (1988).

[20]   M. Endo, T. Tsunoo, and N. Nakamori,"Effect of scatter radiation on image noise in cone beam CT," Proc. SPIE **3977**, 514-521 (2000).

[21]   G. Barnea and C. E. Dick,"Monte Carlo studies of x-ray scattering in transmission diagnostic radiology," Med Phys **13**, 490-495 (1986).

[22] H. P. Chan and K. Doi,"Some properties of photon scattering in water phantoms in diagnostic radiology," Med Phys **13**, 824-830 (1986).

[23] H. P. Chan, Y. Higashida, and K. Doi,"Performance of antiscatter grids in diagnostic radiology: experimental measurements and Monte Carlo simulation studies," Med Phys **12**, 449-454 (1985).

[24] J. M. Boone and J. A. Seibert,"Monte Carlo simulation of the scattered radiation distribution in diagnostic radiology," Med Phys **15**, 713-720 (1988).

[25] J. M. Boone and J. A. Seibert,"An analytical model of the scattered radiation distribution in diagnostic radiology," Med Phys **15**, 721-725 (1988).

[26] H. P. Chan and K. Doi,"Physical characteristics of scattered radiation in diagnostic radiology: Monte Carlo simulation studies," Med Phys **12**, 152-165 (1985).

Physical characterization of a prototype selenium-based full field digital mammography detector

Robert S. Saunders, Jr.
Duke Advanced Imaging Laboratories
Departments of Physics and Radiology, Duke University, Durham, North Carolina 27710

Ehsan Samei
Duke Advanced Imaging Laboratories
Departments of Radiology, Physics, and Biomedical Engineering, Duke University, Durham,
North Carolina 27710

Jonathan L. Jesneck
Duke Advanced Imaging Laboratories
Departments of Biomedical Engineering and Radiology, Duke University, Durham, North Carolina 27710

Joseph Y. Lo
Duke Advanced Imaging Laboratories
Departments of Radiology and Biomedical Engineering, Duke University, Durham, North Carolina 27710

The purpose of this study was to measure experimentally the physical performance of a prototype mammographic imager based on a direct detection, flat-panel array design employing an amorphous selenium converter with 70 μm pixels. The system was characterized for two different anode types, a molybdenum target with molybdenum filtration (Mo/Mo) and a tungsten target with rhodium filtration (W/Rh), at two different energies, 28 and 35 kVp, with approximately 2 mm added aluminum filtration. To measure the resolution, the presampled modulation transfer function (MTF) was measured using an edge method. The normalized noise power spectrum (NNPS) was measured by two-dimensional Fourier analysis of uniformly exposed mammograms. The detective quantum efficiencies (DQEs) were computed from the MTFs, the NNPSs, and theoretical ideal signal to noise ratios. The MTF was found to be close to its ideal limit and reached 0.2 at 11.8 mm^{-1} and 0.1 at 14.1 mm^{-1} for images acquired at an RQA-M2 technique (Mo/Mo anode, 28 kVp, 2 mm Al). Using a tungsten technique (MW2; W/Rh anode, 28 kVp, 2 mm Al), the MTF went to 0.2 at 11.2 mm^{-1} and to 0.1 at 13.3 mm^{-1}. The DQE reached a maximum value of 54% at 1.35 mm^{-1} for the RQA-M2 technique at 1.6 μC/kg and achieved a peak value of 64% at 1.75 mm^{-1} for the tungsten technique (MW2) at 1.9 μC/kg. Nevertheless, the DQE showed strong exposure and frequency dependencies. The results indicated that the detector offered high MTFs and DQEs, but structured noise effects may require improved calibration before clinical implementation. © *2005 American Association of Physicists in Medicine.* [DOI: 10.1118/1.1855033]

Key words: image quality, mammography, modulation transfer function, normalized noise power spectrum, detective quantum efficiency, digital imaging

I. INTRODUCTION

Breast cancer remains the second leading cause of cancer death for women in the United States. The American Cancer Society (ACS) estimates that in 2004, 215 990 new cases of invasive breast cancer will be diagnosed and 40 110 women will die from the disease in the United States.[1] Early detection of this disease holds the key for survival, as more treatment options exist for early stage cancers and treatments tend to be more successful at this stage. X-ray mammography continues to be widely regarded as the most effective early-detection screening tool available today.[2,3] X-ray mammography places severe demands, however, on an imaging system. A system must capture small, low contrast anatomical details, as the early signs of cancer are often very subtle. While mammography has experienced notable advancements in recent years, further improvement is required as up to 22% of cancers are missed at the initial screening.[4]

Full Field Digital Mammography (FFDM) offers the promise of improving mammographic image quality and therefore increasing the utility of this screening procedure.[5–7] As images are stored in a digital format, a radiologist can view the images at any workstation or many clinicians can have simultaneous access to the images. The use of image processing algorithms enhances various features in the image. In addition, these systems have the potential to improve mammographic imaging by separating each stage of the imaging chain, from detection to image processing to display, allowing each step to be independently optimized.

The current state of the art in digital mammography is solid-state flat-panel detectors.[8] Flat-panel detectors can be subdivided into two categories, direct and indirect, named

for the mechanisms used to detect x-rays.[9,10] In direct detectors, a photoconductive layer absorbs an incoming x-ray photon and converts it to electric charge. A voltage applied across the photoconductor then draws the charges toward the pixel electrodes.[11,12] In contrast, indirect detectors utilize a scintillation layer that converts the x-ray photon into visible light photons, which are subsequently absorbed by photosensitive elements.[13,14] Because of the different physical mechanisms used to detect photons, the image quality characteristics of these detectors differ substantially. Several prior studies have substantiated some of these differences.[9,10,15,16] In addition, two previous studies have examined limited aspects of image quality for selected mammographic detectors using amorphous selenium.[17,18]

The main purpose of this work was to comprehensively evaluate the physical image quality characteristics of an early prototype mammographic detector based on a direct detection flat-panel array design that employed an amorphous selenium converter. Three key metrics of image quality were evaluated for several radiographic techniques, the modulation transfer function (MTF), normalized noise power spectrum (NNPS), and detective quantum efficiency (DQE), which described the resolution, noise, and signal to noise performance of the detector, respectively.[19–24] As previous research has shown that selenium detectors can exhibit image lag and ghosting,[25] this research also examined the lag performance of the detector.

A secondary objective of this research was to consider new beam qualities for digital mammography. Traditionally, screen-film mammography was performed using a beam from a molybdenum target with molybdenum filtration.[26] This beam quality might not be optimal for digital mammography, however, given the different energy sensitivities and greater dynamic range of digital detectors. Several researchers had suggested that other beam qualities could allow for better image quality for digital mammography.[17,27–30] Therefore, the study examined the image quality characteristics for two different anode types, a molybdenum target with molybdenum filtration and a tungsten target with rhodium filtration, and for two different energies, 28 kVp and 35 kVp, with added aluminum filtration.

II. METHODS AND MATERIALS

A. Detector description

The detector investigated in this study was an early prototype mammographic imager based on a direct detection flat-panel array design that employed an amorphous selenium converter (Mammomat Novation[DR]; Siemens Medical Solutions; Erlangen, Germany). The detector utilized a 250 µm amorphous selenium photoconductive layer coupled to a matrix of pixels, each with a storage capacitor and amorphous silicon switching transistor.[18] The active detector area was 23.3 cm × 28.7 cm consisting of 3328 × 4096 square pixels. Each pixel was placed with a 70 µm pixel pitch and offered a fill factor of greater than 90%. This product has since received FDA approval.
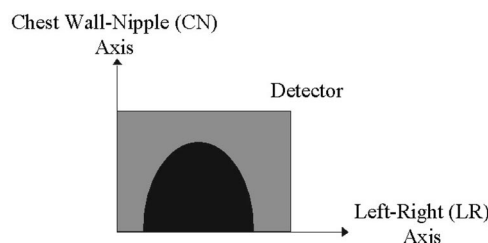

FIG. 1. Coordinate system for physical measurements. These axes are labeled by the anatomy imaged in the craniocaudal view.

Prior to evaluation, the standard antiscatter grid and compression paddle were removed from the system. For most measurements, the standard detector cover was placed on the system. For the MTF measurements, the detector cover was removed so that an edge device could be placed as close as possible to the active selenium layer to minimize focal spot blur.

The coordinate system used to describe the system, as shown in Fig. 1, referred to the anatomical features as viewed on a craniocaudal view. There were two main axes, the chest wall–nipple (CN) axis as well as the left–right (LR) axis. By examining the system performance along these two orthogonal axes, one was able to identify any asymmetries.

B. Image acquisition

A high-frequency, multiphase x-ray generator (Mammomat Novation[DR]), for which the high voltage accuracy was verified to be within ±5%, served as the x-ray source for the system. The anode was operated with a large focal spot of 0.3 mm (IEC), nominal, for all image acquisitions. No postprocessing was applied to the images. All images were transferred to a research computer as 14-bit, raw data for analysis.

Prior to image acquisition, the detector underwent routine detector calibration to correct for dead pixels and gain nonuniformities. The process formed a dead pixel map by detecting inactive pixels in a flat-field image acquired at 28 kVp with a 4 cm PMMA slab in the beam. A gain map was similarly computed from the average of eight flat-field images also acquired at 28 kVp with a 4 cm PMMA slab in the beam. As no images in this research utilized an antiscatter grid, the calibration was performed without a grid in place. The system corrected all subsequently acquired images using the gain and dead pixel maps.

For all image acquisitions, the exposure to the detector was measured using a calibrated ionization chamber (1515 x-ray monitor with 10X5-6M dedicated mammography ionization chamber, Radcal Corporation, Monrovia, CA) placed at 48 cm from the focal spot. As reported in previous studies, this ionization chamber had little energy dependence over mammographic energies.[31] Manufacturer specifications note that the calibration accuracy of the chamber was ±4% (at 20 kVp, 0.26 mm Al HVL) with ±5% energy dependence in the 10 keV to 40 keV range. The exposures incident on the detector, located at 65 cm distance from the focal spot, were estimated from the measured exposure values using the inverse-square law.

Name	Anode target	Anode filtration	kVp	Added filtration (mm Al)	Half-value layer (mm Al)
RQA-M2	Molybdenum	Molybdenum (30 μm)	28	2	0.6
RQA-M4	Molybdenum	Molybdenum (30 μm)	35	1.8	0.68
MW2	Tungsten	Rhodium (50 μm)	28	2	0.79
MW4	Tungsten	Rhodium (50 μm)	35	2	0.92

Four different beam qualities were utilized for the image quality measurements, as outlined in Table I. Two molybdenum techniques, RQA-M2 and RQA-M4, were chosen from the International Electrotechnical Commission (IEC) standard 61267-2.[32] The standard specified the anode type, anode filtration, kVp, and half-value layer for each beam quality. Aluminum filtration was then placed in the beam to produce the desired half-value layer. The IEC standard did not include corresponding tungsten techniques for mammographic applications. To facilitate meaningful comparisons between detector systems, two additional tungsten techniques, MW2 and MW4, were used that had similar characteristics to those of the molybdenum techniques, RQA-M2 and RQA-M4. The half-value layers for these four beam qualities were measured using a narrow geometry and added aluminum filtration in 0.1 mm increments around the estimated half-value layer thickness. The half-value layer thicknesses were then estimated from logarithmic interpolation of the measured exposure values.[32]

C. Linearity

Linearity was determined by exposing the detector to a wide range of uniform x-ray exposures for each of the four radiographic techniques described above. The average pixel values were computed from a 14.3 cm × 14.3 cm region located near the chest wall section of the detector. From this, the relationships between mean pixel value and exposure were ascertained for each technique.

D. Modulation transfer function

An edge method, reported in previous publications,[9,10,33–36] was used to measure the presampled MTF. A 0.1 mm Pt–Ir edge was placed in contact with the detector at 1 cm distance from the chest wall edge of the detector. The device was oriented at a 3°–6° angle with respect to the pixel array. Edge images were then acquired at each of the four radiographic techniques at relatively high exposure values of 16.2 μC/kg (62.6 mR), 15.3 μC/kg (59.2 mR), 9.52 μC/kg (36.9 mR), and 9.75 μC/kg (37.8 mR) for RQA-M2, RQA-M4, MW2, and MW4 techniques, respectively.

A previously reported routine[37] analyzed the edge images in a region around the edge (21.2 mm × 35.8 mm) to determine the presampled modulation transfer function (MTF). In summary, first a double Radon transformation determined the angle of the edge transition with 0.01° accuracy. The edge spread function (ESF) was computed by projecting the image data along lines parallel to the edge transition using bin sizes of 0.1 pixels. To minimize noise, the ESF was smoothed using a modest fourth-order moving polynomial fit and differentiated to form the line spread function (LSF). A Hanning window with 10 mm width was then applied to the LSF to force the tails of the LSF to zero. Finally, the presampled MTF was computed as the normalized Fast Fourier Transform of the LSF.

E. Normalized noise power spectrum

To characterize the system noise, images were acquired of uniform beams of radiation for the different techniques, while the exposure was simultaneously measured with an ionization chamber. The NNPS was then computed from these flat-field images using previously published methods.[36–38] A large region near the chest wall side of the detector, excluding the edges of the image, was used for analysis. This region was segmented into 256 sequential regions of interest (ROIs) of 128 × 128 pixels. A two-dimensional polynomial surface was subtracted from each region of interest (ROI) to minimize background trending and a Hamming window was applied to each ROI so that the edges of the ROI went to zero. To account for intensity variations in the image, each ROI was then scaled by the ratio of its mean to the mean pixel value of the ROI in the top-left-hand corner of the image. Each ROI was transformed by a two-dimensional FFT and the absolute magnitude squared of each FFT was averaged together to obtain the NNPS. This procedure could be summarized in the following equation:[37]

$$
\text{NNPS}(u,v) = \frac{dA}{M \cdot N^2} \sum_{i=1}^{M} \left\{ \frac{\langle \text{ROI}_i \rangle}{\langle \text{ROI}_1 \rangle} \left| \text{FFT}\left[\frac{1}{\langle \text{ROI}_i \rangle}(\text{ROI}_i - \langle \text{ROI}_i \rangle) \right] \right|^2 \right\},
$$
(1)

where dA represented the pixel area, M described the number of regions of interest in which the image was segmented, N corresponded to the number of pixels along one edge of an ROI, ROI_i referred to a particular region of interest within the flat field image, ROI_1 corresponded to the ROI in the top-left corner of the image, and $\langle \text{ROI}_i \rangle$ was the mean of ROI_i. To summarize this two-dimensional information in one-dimensional form, horizontal and vertical traces were obtained by averaging together the central frequency bands (the central axis and ±5 frequency lines). Radial traces were also obtained by radial averaging.

The magnitude of the NNPS could be related to the image variance using Parseval's Theorem[39] and applying ergodic assumptions. This allowed the replacement of $\langle \text{ROI}_i \rangle$ by $\langle I \rangle$, the mean of the entire image, and the mean variance of the ROIs became the variance of the image, σ^2. One could then show that

TABLE II. Ideal SNR^2/mR values calculated for an energy-integrating detector. The beams were modeled with the specified intrinsic filtrations as well as the experimentally measured half-value layer.

Name	Anode target	Anode filtration	$q_{Ideal}(mm^{-2} mR^{-1})$
RQA-M2	Molybdenum	Molybdenum (30 μm)	46052
RQA-M4	Molybdenum	Molybdenum (30 μm)	52542
MW2	Tungsten	Rhodium (50 μm)	54773
MW4	Tungsten	Rhodium (50 μm)	67781

$$\sum_{u,v} NNPS(u,v) = \frac{N^2 dA}{\langle I \rangle^2} \sigma^2. \tag{2}$$

For a linear, quantum-limited detector, $\langle I \rangle$ and σ^2 are proportional to the exposure, E, which would make the product of the NNPS and exposure independent of exposure. The product of NNPS and exposure was then used as a way to assess how well the detector approximated a quantum-limited detector.

A second examination of system noise utilized a background subtraction method, which isolated the quantum noise components of total system noise.[39–42] An average image was created from ten repeated images acquired with the RQA-M2 technique at 125 mAs. The average image was then subtracted from one of the individual images to form a "background-free" image. The NNPS was then computed from the "background-free" image. To correct for the change in image variance caused by the averaging technique, the NNPS for the "background-free" image was multiplied by $N/(N-1)$, where N equaled 10, the number of images used to create the average image.[42]

F. Detective quantum efficiency

The measured MTF and NNPS were combined to determine the Detective Quantum Efficiency (DQE) as

$$DQE(u) = \frac{MTF^2(u)}{q_{Ideal} \cdot E \cdot NNPS(u)}, \tag{3}$$

where q_{Ideal} described the ideal signal to noise (SNR) ratio squared per unit exposure for an energy-integrating detector, and E represented the exposure value at the detector.[39,43] An x-ray simulation program (xSpect, Henry Ford Health System) was used to calculate the q_{Ideal} using a semiempirical model for the x-ray spectra[44] and the attenuation properties of the material.[33] The q values are reported in Table II.

G. Image lag measurement

The magnitude of the multiplicative image lag was characterized using the procedure described in IEC standard 62220-1.[45] First, an image was acquired of a uniform radiation field at a given exposure at time t_1. A second image was then acquired of an edge device at the same exposure level at time t_2. After a specified delay time τ, a third image was acquired of a uniform radiation field at time t_3. This procedure then measured the residual signal from the edge device in the later image.
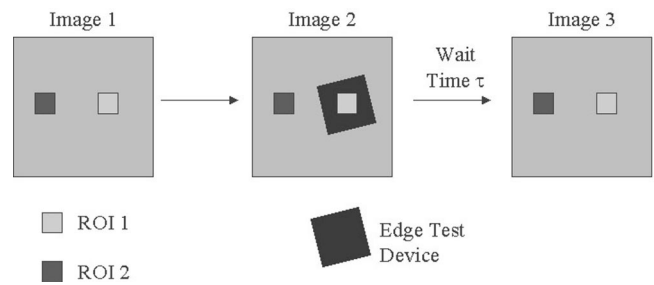
FIG. 2. Illustration of the lag measurement procedure as described in the IEC standard 62220-1.

To determine the magnitude of the residual signal, the image data were examined for two regions within all three images, as shown graphically in Fig. 2. An ROI, ROI_1, was placed in an area of image 2 that contained the edge device. A second ROI, ROI_2, was placed in an area of image 2 that was outside of the edge device. The detector was judged to have negligible residual signal with time delay τ if it passed the following test:[45]
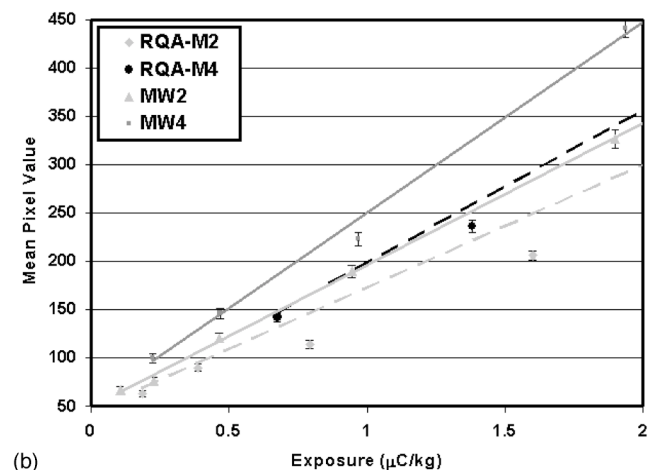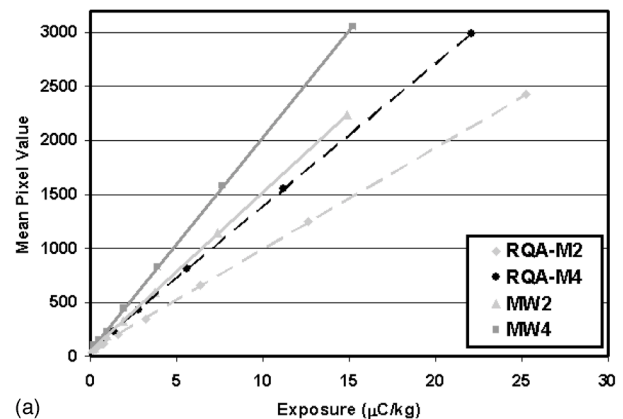

FIG. 3. Plot of mean pixel value versus exposure for two Mo/Mo beams and two W/Rh beams over the (a) entire measured exposure range and (b) the lower exposure range. While the detector exhibits good linearity over the entire range, divergences from linearity occur in the lower exposure range.
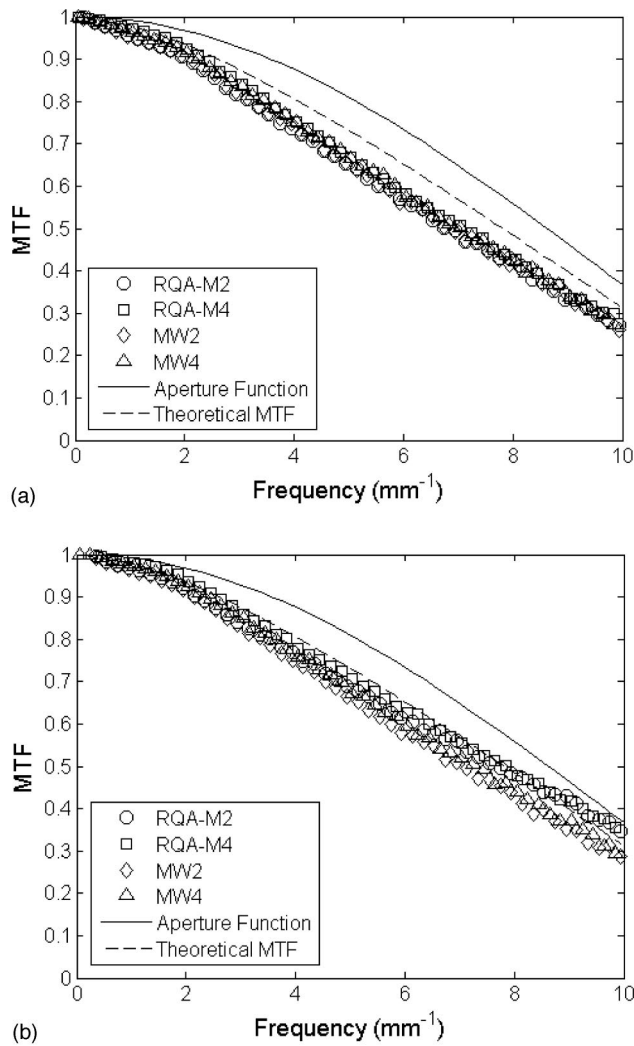
(a)

(b)

FIG. 4. Plot of detector MTF along (a) CN and (b) LR axes for two Mo/Mo beams and two W/Rh beams. The MTFs for the four beams are very similar for the CN axis, but differ along the LR axis. The pixel aperture limit and theoretical MTF (Ref. 18) are included for reference.

$$\frac{\left|(\zeta_{t_1} - \eta_{t_1}) - (\zeta_{t_3} - \eta_{t_3})\right|}{\dfrac{\eta_{t_1} + \eta_{t_3}}{2}} \leq 0.005, \tag{4}$$

where ζ_t and η_t represented the mean of ROI$_1$ and ROI$_2$ at time t, respectively. The IEC chose the threshold of 0.005 as the maximum allowable level of residual signal.

III. RESULTS

Figure 3 illustrates the relationship between pixel value and exposure for the detector. In general, the system showed a very linear response with correlation coefficients for linear regression fits greater than 0.999. One interesting trend was that the detector was slightly more sensitive to the W/Rh beam qualities than the Mo/Mo beams, as the tungsten curves resulted in higher slopes and higher pixel values for equivalent exposures. Another trend was revealed by exam-

TABLE III. Summary of the detector MTF properties along CN and LR axes for (top) Mo/Mo beams and (bottom) W/Rh beams. Shown are frequencies at specific MTFs and the MTF at specific frequencies. The MTF for the Mo/Mo beams differed between the CN and LR axes, but was similar for the W/Rh beams.

	RQA-M2 (CN Axis)	RQA-M2 (LR Axis)	RQA-M4 (CN Axis)	RQA-M4 (LR Axis)
0.2 MTF	11.1 mm^{-1}	12.7 mm^{-1}	11.2 mm^{-1}	12.5 mm^{-1}
0.1 MTF	12.8 mm^{-1}	14.8 mm^{-1}	12.8 mm^{-1}	14.5 mm^{-1}
0.5 mm^{-1}	0.98	0.99	0.99	1.0
2.5 mm^{-1}	0.86	0.88	0.89	0.90
5.0 mm^{-1}	0.65	0.70	0.67	0.71
	MW2 (CN Axis)	MW2 (LR Axis)	MW4 (CN Axis)	MW4 (LR Axis)
0.2 MTF	11.1 mm^{-1}	11.4 mm^{-1}	11.2 mm^{-1}	11.5 mm^{-1}
0.1 MTF	12.9 mm^{-1}	13.2 mm^{-1}	12.9 mm^{-1}	13.3 mm^{-1}
0.5 mm^{-1}	0.98	0.99	0.99	0.99
2.5 mm^{-1}	0.86	0.87	0.88	0.89
5.0 mm^{-1}	0.65	0.67	0.67	0.68

ining the lower exposure range, such as that shown in Fig. 3(b), where some deviations from linearity were seen.

To verify image repeatability over time, an ensemble of images was acquired at identical mAs. The mean signal for each image was computed as the average pixel value over a region of interest. These images showed very similar signal levels over time, as the mean signal varied by 0.009% over the entire ensemble of images. In contrast, the spatial deviation, which described how the pixel values varied across each image, reached 3.4% for the lowest exposure images.

The resolution properties of the detector, as represented by the MTF, are shown in Fig. 4 and summarized in Table III. Figure 4(a) illustrates the MTF along the CN axis, while Fig. 4(b) displays the MTF along the LR axis. While the MTFs for the tungsten and molybdenum techniques overlapped considerably for the CN direction, the molybdenum
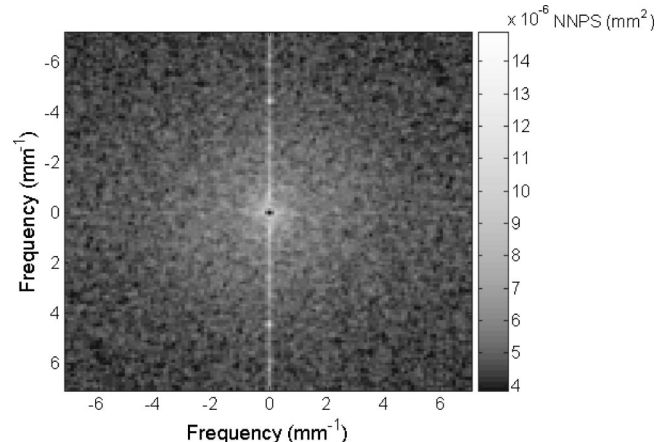

FIG. 5. Two-dimensional NNPS for RQA-M2 beam quality at 1.58 μC/kg exposure. The image is shown in a logarithmic scale. Nonstochastic noise is observed in a frequency band along the CN frequency axis.
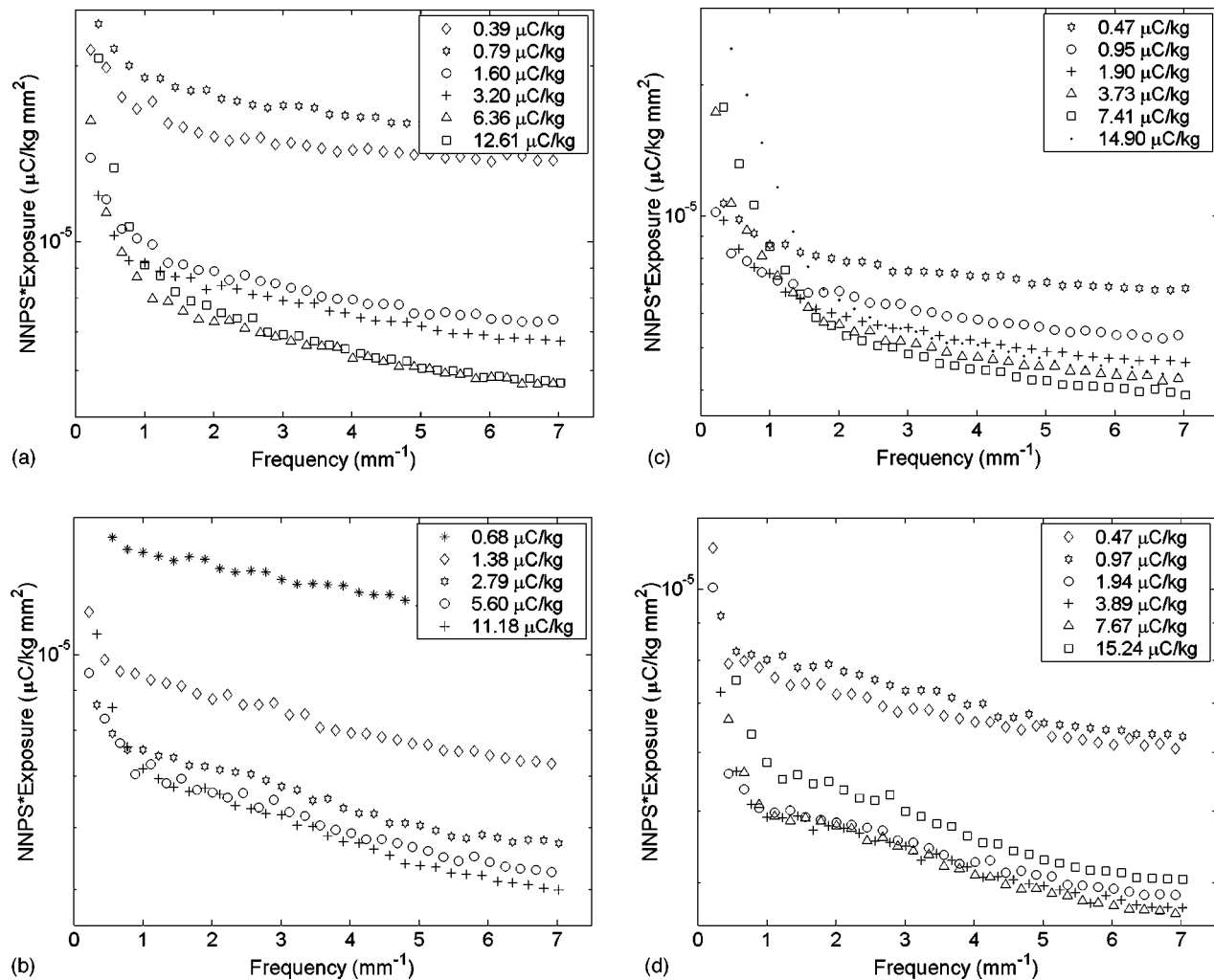
FIG. 6. Radial traces of NNPS multiplied by exposure for (a) RQA-M2, (b) RQA-M4, (c) MW2, and (d) MW4 beam qualities.

and tungsten curves diverged for the LR axis. For reference, Fig. 4 also displays the pixel aperture function and the theoretical limit calculated by Yorker *et al.*[18]

Figure 5 shows an example of a two-dimensional NNPS displayed in a logarithmic scale (RQA-M2 technique, 1.58 μC/kg). The figure demonstrates nonstochastic noise in the CN direction along a band of 0.112 mm^{-1} in width. Similar behavior was observed for other exposures and beam qualities. Figure 6 illustrates the radial NNPS multiplied by exposure. As discussed in Sec. II E, the product of NNPS and exposure should remain constant for strictly quantum-limited detectors, however, the results showed notable exposure dependencies. For lower exposures, the magnitude of this metric decreased to some minimum value as one increased exposure. For several techniques, the magnitude of the metric increased at higher exposures.

Figure 7 illustrates the NNPS calculated through the background subtraction method. The background subtraction method noticeably reduced the low-frequency noise. In addition, the overall magnitude of the NNPS decreased. One
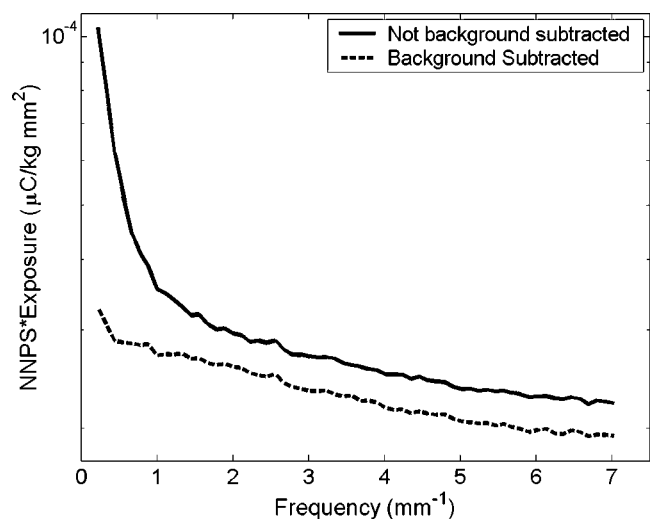

FIG. 7. Radial traces of NNPS multiplied by exposure obtained with and without background subtraction method. The NNPS was obtained using RQA-M2 technique at 12.6 μC/kg. The background subtraction routine reduced the low-frequency noise and lowered overall noise.
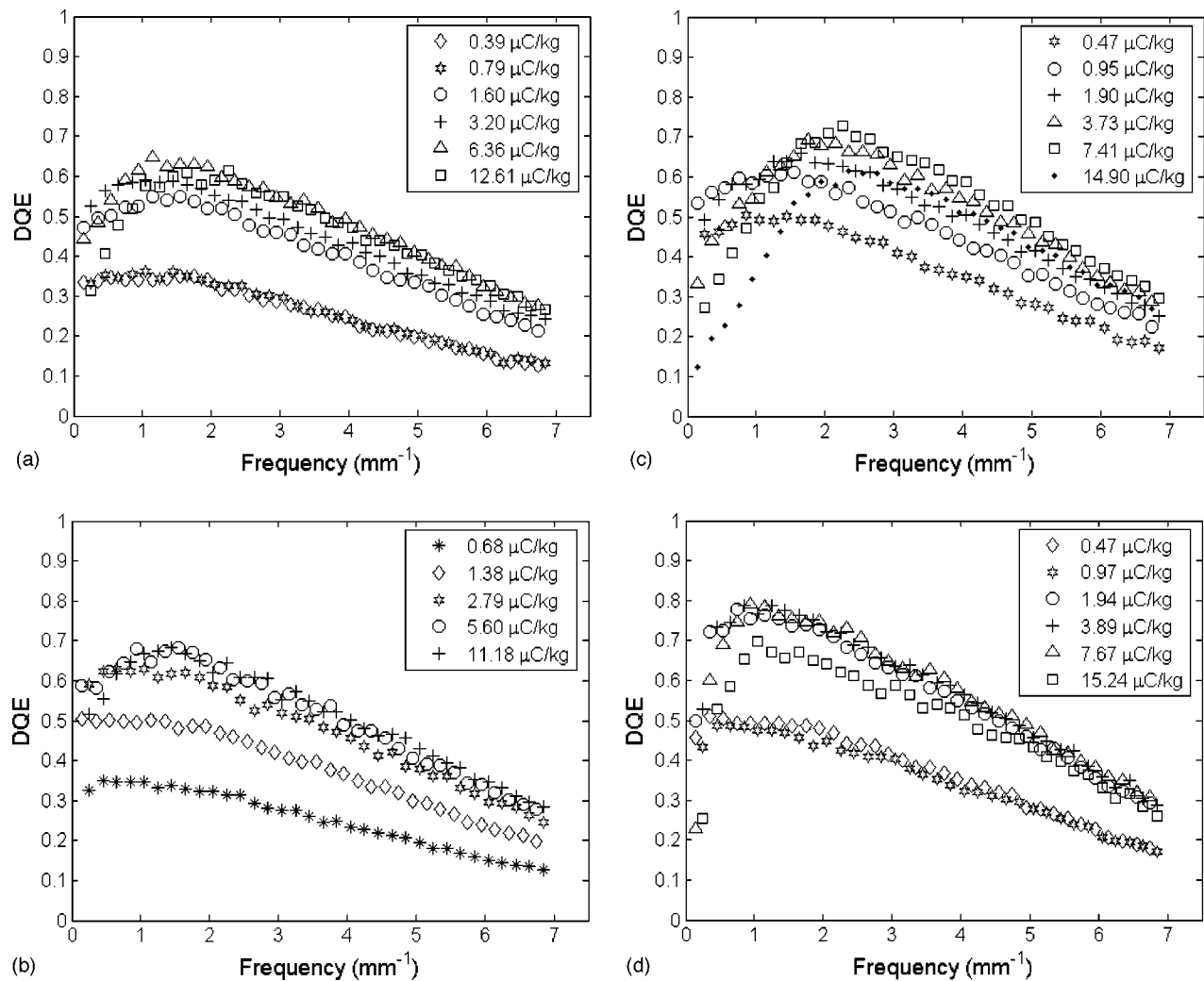
FIG. 8. DQE averaged over CN and LR axes for (a) RQA-M2, (b) RQA-M4, (c) MW2, and (d) MW4 beam qualities.

TABLE IV. Detector DQE along the (top) CN axis and (bottom) LR axes. As the low-frequency noise caused peaks in the DQE, the table reports the maximum DQE value and the frequency at which this maximum occurs.

	RQA-M2 (1.60 μC/kg)	RQA-M4 (2.79 μC/kg)	MW2 (1.90 μC/kg)	MW4 (1.94 μC/kg)	Background subtracted RQA-M2 (12.6 μC/kg)
0.15 mm^{-1}	46%	59%	46%	50%	73%
2.5 mm^{-1}	49%	55%	61%	66%	64%
5.0 mm^{-1}	31%	36%	41%	44%	44%
Peak	55%	63%	66%	77%	73%
	1.25 mm^{-1}	0.85 mm^{-1}	1.55 mm^{-1}	0.85 mm^{-1}	0.15 mm^{-1}

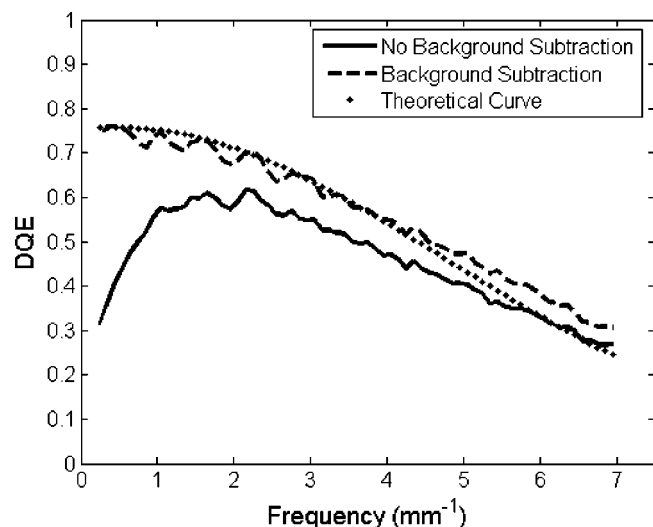|	RQA-M2 (1.60 μC/kg)	RQA-M4 (2.79 μC/kg)	MW2 (1.90 μC/kg)	MW4 (1.94 μC/kg)	Background subtracted RQA-M2 (12.6 μC/kg)
0.15 mm^{-1}	47%	59%	47%	52%	77%
2.5 mm^{-1}	50%	57%	61%	70%	68%
5.0 mm^{-1}	34%	41%	41%	47%	49%
Peak	53%	63%	63%	77%	77%
	1.45 mm^{-1}	0.95 mm^{-1}	1.85 mm^{-1}	0.85 mm	0.15 mm^{-1}

FIG. 9. DQE calculated using background subtraction method averaged over the CN and LR axes. The DQE was computed for RQA-M2 technique at 12.6 μC/kg. The background subtraction routine reduced the low frequency peaking. The plot also shows a theoretical estimation of the DQE (Ref. 46) for a similar detector (200 μm selenium layer, 85 μm pixel size).

TABLE V. Lag properties of the detector. The lag tests were executed in the order shown in the table, with the top three rows measuring signal retention after a 3 min decay time, then a gap of 10 mins, and the bottom three rows measured signal retention after a 5 min delay time. The metric corresponded to IEC 62220-1 with values less than 0.005 acceptable under the IEC guidelines.

Test Number	Exposure (μGy)	Decay time (min)	Metric	Acceptable residual signal?
1	75	3	0.002	Yes
2	150	3	0.0048	Yes
3	200	3	0.0044	Yes
		Ten minute wait		
4	75	5	0.047	No
5	150	5	0.013	No
6	200	5	0.0022	Yes

should note that the noise was corrected for the change in noise variance due to the subtraction technique, as discussed in Sec. II E.

Figure 8 shows the DQE measured for all techniques in the axial (the average of CN and LR axes) direction. The DQE curves showed low frequency peaking, in that the DQE exhibited a sharp increase at lower frequencies. The strong low-frequency component of the NNPS led to this unusual behavior. Moreover, the DQE increased with exposure for lower exposure values, reached a peak value, and then decreased for higher exposures. This was also expected from the behavior of the NNPS. The DQE is summarized in Table IV for all four techniques along both CN and LR axes.

To separate the fixed pattern noise from quantum noise effects, the DQE was calculated with the background subtraction method. Figure 9 illustrates the DQE computed with this method in the axial direction. By eliminating the fixed pattern noise, the low-frequency peaking in the DQE was removed and the overall efficiency increased. This figure also includes the theoretical DQE calculated for a similar detector for reference (200 μm selenium layer, 85 μm pixel size).[46]

The results from lag measurements are summarized in Table V. In general, the image lag for the detector passed the test established by IEC 62220-1. An interesting phenomenon occurred for the fourth test (75 μGy exposure, 5 min decay time). The residual signal level was unacceptably high for this test, even though a similar test (75 μGy exposure, 3 min decay time) produced acceptable levels of residual signal.

IV. DISCUSSION

Digital mammography has begun to replace screen-film systems in some clinical settings. The motivation for this change includes several logistical considerations, such as convenient archiving and display, and potential image quality advantages. The two flat-panel technologies currently offered, direct and indirect, vary markedly in terms of their image quality characteristics. Direct detectors tend to enjoy higher resolution than indirect detectors. However, they are often less efficient than their indirect counterparts.[9,10] In this study, we evaluated all physical properties of a particular direct flat-panel detector, including resolution, noise, and efficiency, to enable a thorough comparison between that detector and others.

Several other investigators have examined the physical characteristics of flat-panel mammographic imagers. As such, the results from this system characterization must be reported in the context of the performance of other systems. When considering previous measurements, one should note any differences in beam energies and filtrations. Most prior studies utilized molybdenum anodes with molybdenum filtration at 28 kVp, but often utilized a breast equivalent phantom for further filtration.[18,46–49] While this should still allow for reasonable comparisons between MTFs, these differences would make comparisons of DQE curves more challenging.

Compared to previous measurements of indirect flat-panel imagers,[47,48] the current system exhibited a higher MTF. At low frequencies, our MTF was similar to other direct flat-panel imagers, but our MTF was higher at higher frequencies.[46,49] As the pixel size served as the primary limiter of the resolution of a direct detector, with some blurring effects from backscatter and reabsorption of K x-rays, the similarity between direct detectors was reasonable. At similar exposures, the DQE of the system was generally higher than that of indirect flat-panel imagers, although the low-frequency peaking complicated this comparison.[47] In comparison to the work by Jee, the High Light (HL) output configuration produced a higher DQE but the High Resolution (HR) configuration appeared to produce a lower DQE than our current system.[48] The direct detector evaluated by Zhao produced a generally higher DQE, with constant behavior for different exposures, although there were significant differences between the axes.[46] One interesting result was that
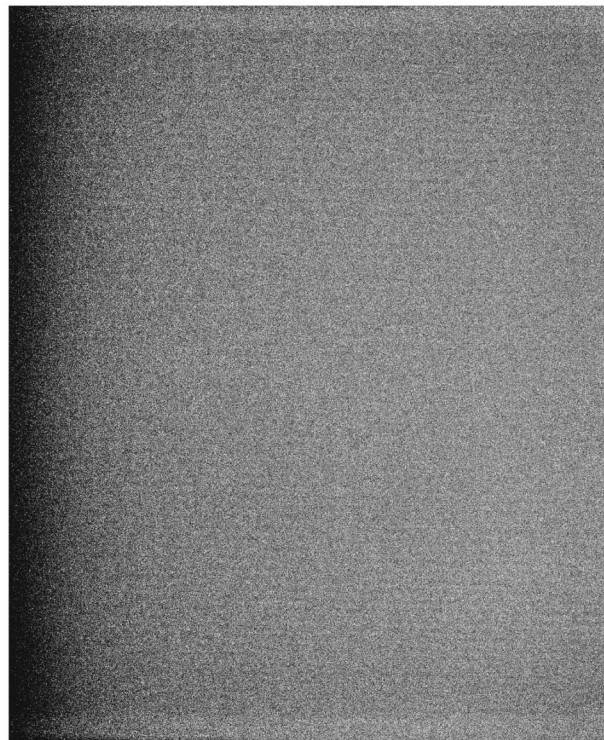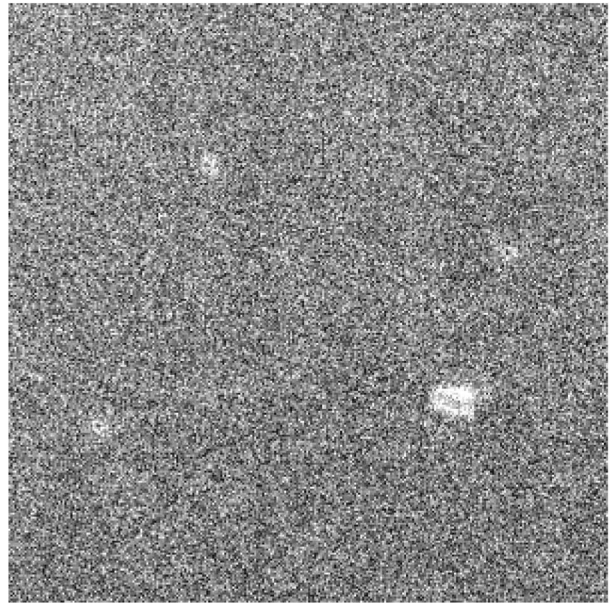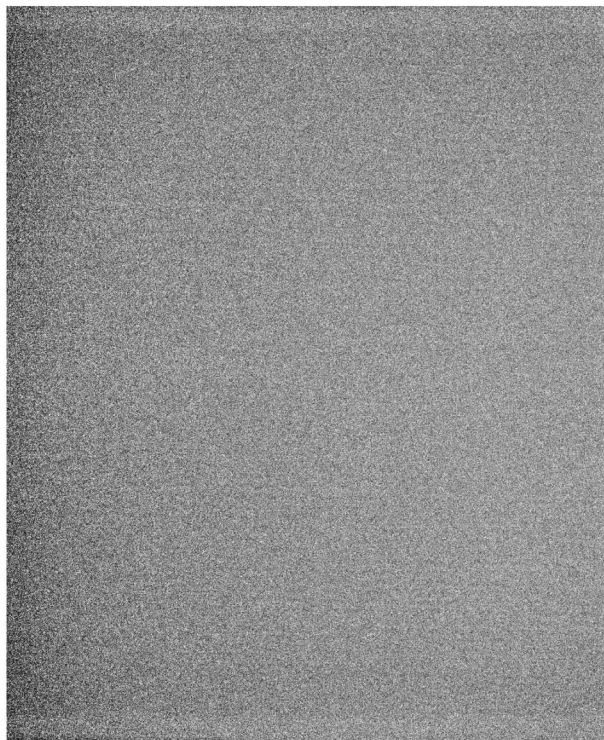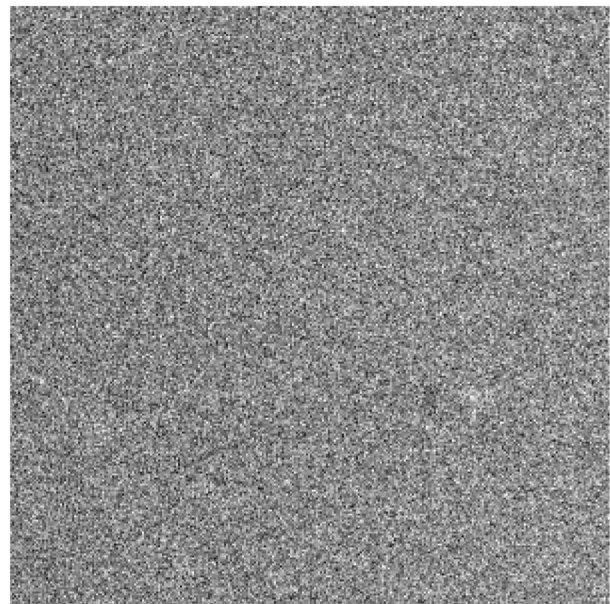
FIG. 10. Example image of uniform beam of radiation (a) before and (b) after the application of a secondary gain correction from the average of 10 images. The larger images (physical size: 23.3 cm $\times$ 28.7 cm) show the differences in large-scale gain nonuniformities. Zoomed portion of the images (2.1 cm $\times$ 2.1 cm) highlighting pixel artifacts (c) before and (d) after gain calibration. The gain calibration largely removes the pixel artifacts from the individual images.
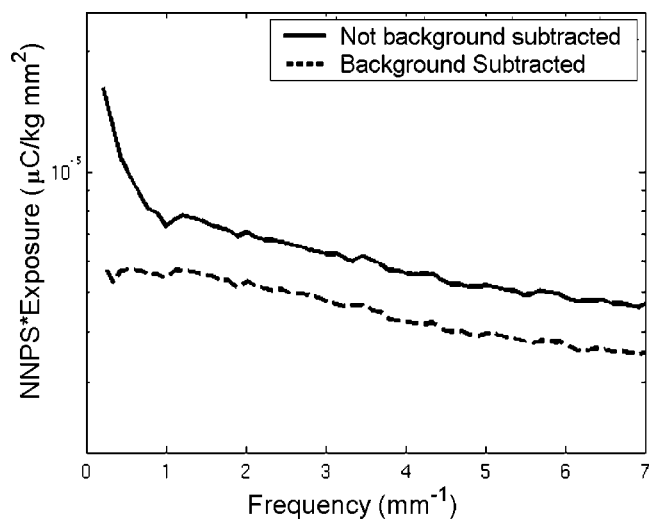
FIG. 11. Radial traces of NNPS multiplied by exposure obtained with and without background subtraction method for a second prototype detector. The NNPS was obtained using RQA-M2 technique at 7.67 μC/kg. As with the prototype system, this detector unit also exhibits significant stochastic noise.

some previous studies on selenium detector systems[46,49] have also observed some low-frequency peaking in the DQE, although it was less pronounced than that observed in this study.

Yorker *et al.* have published the MTF and DQE measurements of a similar mammographic detector with an identical pixel size.[18] That study examined the MTF and DQE for one radiographic technique using a molybdenum anode operated at 28 kVp with molybdenum filtration. The reported MTF for this technique was very similar to our measured MTF acquired at the RQA-M2 technique. At similar exposures, our DQE acquired at RQA-M2 also appeared comparable to that of Yorker *et al.* However, given the fact that Yorker *et al.* used a 4.2 cm breast phantom filter in the beam, quantitative comparisons are not straightforward.

Several researchers have explored the theoretical properties of selenium-based flat-panel imagers operated at mammographic energies. The properties of this system compared favorably with these theoretical calculations. As shown in Figs. 4(a) and 4(b), the MTF of this system remained close to the theoretical limit, as calculated by Yorker *et al.*[18] When using the background subtraction method to remove fixed pattern noise, the DQE of the system appeared similar to its theoretical value for a similar detector.[46] The difference between the theoretical and experimental values were likely due to the assumptions behind the theoretical calculation, which assumed a 200 μm selenium layer and 85 μm pixel size. These theoretical calculations should underestimate the actual detector efficiency, as a larger selenium layer will more efficiently capture x-ray photons and a smaller pixel size should boost the higher frequency portions of the DQE. Notwithstanding, the experimental results for the MTF and DQE largely agreed with their theoretical values.

This prototype detector had very favorable resolution properties, as shown by its MTF. There was an asymmetry in the MTF, however, as the tungsten and molybdenum curves

overlapped for the CN axis but diverged for the LR axis. The difference might be attributed to the differences between the focal spots for the two anodes, in terms of both shape and location. The impact of focal spot blur should be minimal as the edge was placed directly on the detector. Further work remains needed to evaluate the focal spot properties for both anode varieties and to determine whether this phenomenon occurs with different tubes.

The prototype detector did show significant structured noise contributions. This could be decomposed into two factors: (a) low frequency trending over the image and (b) pixel artifacts. The trending was expressed as a strong low-frequency component of the NNPS. In contrast, pixel artifacts were similar to delta functions and elevated all frequencies of the NNPS. A background subtraction method eliminated both of these factors, so one was unable to determine the relative magnitude of either individually. Therefore, when one compared the NNPS calculated using the background-subtraction technique to that calculated using standard techniques, one noticed a decrease in the low-frequency noise contribution as well as an overall decrease in the magnitude of the NNPS. This was reflected in the DQE as well. When the DQE was calculated using background-subtraction techniques, the sharp low-frequency drop was eliminated and the overall curve was shifted upwards because of the decrease in noise.

Several of these noise concerns could be mitigated by additional gain calibration after the gain calibration performed by the system. To examine the benefits of further gain calibration, a gain map was created by averaging ten uniform images together. This gain map was then applied to a subsequently acquired image. The effect of the gain calibration is shown in Fig. 10 and displayed with identical window and level settings. The prominent trending was greatly diminished and many of the pixel artifacts were eliminated. To assess whether the problem observed was unique to the prototype detector tested, a follow-up experiment was conducted on a more recent prototype device to learn whether it exhibited noise properties similar to the earlier prototype. This experiment compared the NNPS calculated with and without the background subtraction methodology, as shown in Fig. 11. The background subtraction proved to similarly remove significant nonuniformities, which indicated that the images after system gain calibration retained substantial structured noise in the second prototype as well.

The detector was evaluated for four different beam qualities. Two beams used a molybdenum anode with molybdenum filtration and two used a tungsten anode with rhodium filtration. The tungsten beam qualities were developed specifically for this study and inspired by IEC standards. The detector appeared to be slightly more sensitive to the tungsten beams, as shown in the exposure-pixel value relationship. Moreover, the DQEs for the tungsten beams were higher than those for the molybdenum beams, although this was obscured by the peaking in the DQE curves. This suggests that tungsten beams might produce higher quality images with digital detectors than the traditional molybdenum beams.

Using the parameters established by IEC 62220-1, the image lag appeared within reasonable parameters. Nevertheless, high exposures led to unusual behavior in signal retention, affecting other exposures even after a significant length of time. Ten minutes before the fourth lag test (75 μGy exposure, 5 min decay time), a 200 μGy lag test was conducted. It appeared that this high exposure still affected the detector after 10 mins, as a 75 μGy exposure should not have higher residual signal after a 5 min decay time than it would after a 3 min decay time. The mechanisms for this behavior are unknown and suggest additional investigation into signal retention properties of selenium.

V. CONCLUSIONS

This study reported an assessment of image quality for a prototype mammographic imager based on a direct-detection, flat-panel array employing an amorphous selenium converter. The results indicated that the detector had strong potential for capturing high-frequency information, as exhibited by its high MTF. In addition, the DQE of the detector approached the high value of 75%–80%. Yet, suboptimal calibration affected the DQE performance of the system, underscoring the importance of careful gain and dead pixel corrections in reducing detector nonuniformities. Finally, this study introduced two new radiographic techniques utilizing tungsten anodes for the assessment of mammographic systems, which will facilitate the future comparisons of detector characteristics operated with tungsten anodes.

ACKNOWLEDGMENTS

[1]American Cancer Society, *Cancer Facts and Figures 2004* (American Cancer Society, Atlanta, GA, 2004).

[2]I. Jatoi, "Breast cancer screening," Am. J. Surg. **177**, 518–524 (1999).

[3]K. Kerlikowske, D. Grady, S. M. Rubin, C. Sandrock, and V. L. Ernster, "Efficacy of screening mammography: A meta-analysis," J. Am. Med. Assoc. **273**, 149–154 (1995).

[4]R. E. Bird, T. W. Wallace, and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology **184**, 613-617 (1992).

[5]E. D. Pisano, "Current status of full-field digital mammography," Radiology **214**, 26–28 (2000).

[6]L. Fajardo and M. Williams, "The clinical potential of digital mammography," in *Digital Mammography '96*, edited by K. Doi, M. Giger, R. Nishikawa, and R. Schmidt (Elsevier, New York, 1996), pp. 43–52.

[7]A. Maidment and M. Yaffee, "Scanned-slot digital mammography," Proc. SPIE **1231**, 316–326 (1990).

[8]M. J. Yaffe and J. A. Rowlands, "X-ray detectors for digital radiography," Phys. Med. Biol. **42**, 1–39 (1997).

[9]E. Samei and M. J. Flynn, "An experimental comparison of detector performance for direct and indirect digital radiography systems," Med. Phys. **30**, 608–622 (2003).

[10]E. Samei, M. J. Flynn, H. G. Chotas, and J. T. Dobbins, III, "DQE of direct and indirect digital radiographic systems," Proc. SPIE **4320**, 189–197 (2001).

[11]W. Zhao, I. Blevis, S. Germann, J. A. Rowlands, D. Waechter, and Z. S. Huang, "Digital radiology using active matrix readout of amorphous selenium: Construction and evaluation of a prototype real-time detector," Med. Phys. **24**, 1834–1843 (1997).

[12]D. L. Lee, L. K. Cheung, B. Rodricks, and G. F. Powell, "Improved imaging performance of a 14×17-inch direct radiography [trademark] system using Se/TFT detector," Proc. SPIE **3336**, 14–23 (1998).

[13]J. A. Rowlands and J. Yorkston, "Flat panel detectors for digital radiography," in *Handbook of Medical Imaging*, edited by H. K. J. Beutel, H. L. Kundel, and R. Van Metter (SPIE, Washington, D.C., 2000), Vol. 1, pp. 223–328.

[14]J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, and I. A. Cunningham, "Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology," Med. Phys. **25**, 614–628 (1998).

[15]Y. El-Mohri, L. E. Antonuk, K.-W. Jee, Y. Kang, Y. Li, A. Sawant, Z. Su, Y. Wang, J. Yamamoto, and Q. Zhao, "Evaluation of novel direct- and indirect-detection active matrix flat-panel imagers (AMFPIs) for mammography," Proc. SPIE **5030**, 168–180 (2003).

[16]K.-W. Jee, L. E. Antonuk and Y. El-Mohri, "Evaluation of direct-detection and indirect-detection active matrix, flat-panel imagers (AMF-PIs) for digital mammography," Proc. SPIE **4320**, 13–23 (2001).

[17]R. S. Saunders Jr., E. Samei, J. Y. Lo, and J. L. Jesneck, "Physical characterization of a selenium-based full field digital mammography detector," *7th International Workshop on Digital Mammography*, 2004.

[18]J. G. Yorker, L. S. Jeromin, D. L. Y. Lee, E. F. Palecki, K. P. Golden, and Z. Jing, "Characterization of a full-field digital mammography detector based on direct x-ray conversion in selenium," Proc. SPIE **4682**, 21–29 (2002).

[19]H. Fujita, K. Doi, and M. L. Giger, "Investigation of basic imaging properties in digital radiography 6. MTFs of II-TV digital imaging-systems," Med. Phys. **12**, 713–720 (1985).

[20]M. Spahn, M. Strotzer, M. Volk, S. Bohm, B. Geiger, G. Hahm, and S. Feuerbach, "Digital radiography with a large-area, amorphous-silicon, flat-panel x-ray detector system," Invest. Radiol. **35**, 260–266 (2000).

[21]C. D. Bradford, W. W. Peppler, and J. M. Waidelich, "Use of a slit camera for MTF measurements," Med. Phys. **26**, 2286–2294 (1999).

[22]M. B. Williams, P. A. Mangiafico, and P. U. Simoni, "Noise power spectra of images from digital mammography detectors," Med. Phys. **26**, 1279–1293 (1999).

[23]A. D. A. Maidment and M. J. Yaffe, "Analysis of the spatial-frequency-dependent DQE of optically coupled digital mammography detectors," Med. Phys. **21**, 721–729 (1994).

[24]A. Workman and D. S. Brettle, "Physical performance measures of radiographic imaging systems," Dentomaxillofac Radiol. **26**, 139–146 (1997).

[25]W. Zhao, G. DeCrescenzo, and J. A. Rowlands, "Investigation of lag and ghosting in amorphous selenium flat-panel x-ray detectors," Proc. SPIE **4682**, 9–20 (2002).

[26]M. Yaffe, "Digital mammography," in *Handbook of Medical Imaging*, edited by H. K. J. Beutel and R. V. Metter (SPIE, Washington, D.C., 2000), Vol. 1, pp. 329–372.

[27]R. Fahrig and M. J. Yaffe, "Optimization of spectral shape in digital mammography: Dependence on anode material, breast thickness, and lesion type," Med. Phys. **21**, 1473–1481 (1994).

[28]R. L. McKinley, M. P. Tornai, E. Samei, and M. L. Bradshaw, "Simulation study of a quasi-monochromatic beam for x-ray computed mammotomography," Med. Phys. **31**, 800–813 (2004).

[29]E. A. Berns, R. E. Hendrick, and G. R. Cutter, "Optimization of technique factors for a silicon diode array full-field digital mammography system and comparison to screen-film mammography with matched average glandular dose," Med. Phys. **30**, 334–340 (2003).

[30]D. R. Dance, A. K. Thilander, M. Sandborg, C. L. Skinner, I. A. Castellano, and G. A. Carlsson, "Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: A Monte Carlo study," Br. J. Radiol. **73**, 1056–1067 (2000).

[31]L. K. Wagner, B. R. Archer, and F. Cerra, "On the measurement of half-value layer in film-screen mammography," Med. Phys. **17**, 989–997 (1990).

[32]International Electrotechnical Commission, Medical diagnostic x-ray equipment-Radiation conditions for use in the determination of characteristics (IEC-61267, Geneva, Switzerland, 2003).

[33]E. Samei and M. J. Flynn, "An experimental comparison of detector performance for computed radiography systems," Med. Phys. **29**, 447–459 (2002).

[34]E. Samei, "Image quality in two phosphor-based flat panel digital radiographic detectors," Med. Phys. **30**, 1747–1757 (2003).

[35]E. Samei, M. J. Flynn, and D. A. Reimann, "A method for measuring the presampled MTF of digital radiographic systems using an edge test device," Med. Phys. **25**, 102–113 (1998).

[36]M. J. Flynn and E. Samei, "Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems," Med. Phys. **26**, 1612–1623 (1999).

[37]R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med. Phys. **30**, 3006–3017 (2003).

[38]M. J. Flynn, E. Samei, and D. A. Reimann, "Experimental comparison of noise and resolution for 2K and 4K storage phosphor chest radiography systems," Radiology **201**, 1082–1082 (1996).

[39]J. Dobbins, "Image quality metrics for digital systems," in *Handbook of Medical Imaging*, edited by H. K. J. Beutel and R. V. Metter (SPIE, Washington, D.C., 2000), Vol. 1, pp. 163–222.

[40]M. L. Giger, K. Doi, and C. E. Metz, "Investigation of basic imaging properties in digital radiography 2. Noise wiener spectrum," Med. Phys. **11**, 797–805 (1984).

[41]P. R. Granfors and R. Aufrichtig, "Performance of a 41×41-cm2 amorphous silicon flat panel x-ray detector for radiographic imaging applications," Med. Phys. **27**, 1324–1331 (2000).

[42]R. J. Warp and J. T. Dobbins, "Quantitative evaluation of noise reduction strategies in dual-energy imaging," Med. Phys. **30**, 190–198 (2003).

[43]J. C. Dainty and R. Shaw, *Image Science: Principles, Analysis and Evaluation of Photographic-Type Imaging Processes* (Academic, New York, 1974).

[44]E. Storm, "Calculated bremsstrahlung spectra from thick tungsten targets," Phys. Rev. APhys. Rev. A **5**, 2328 (1972).

[45]International Electrotechnical Commission, Medical electrical equipment — Characteristics of digital x-ray imaging devices — Part 1: Determination of the detective quantum efficiency (IEC 62220-1, Geneva, Switzerland, 2003).

[46]W. Zhao, W. G. Ji, A. Debrie, and J. A. Rowlands, "Imaging performance of amorphous selenium based flat-panel detectors for digital mammography: Characterization of a small area prototype detector," Med. Phys. **30**, 254–263 (2003).

[47]S. Vedantham, A. Karellas, S. Suryanarayanan, D. Albagli, S. Han, E. J. Tkaczyk, C. E. Landberg, B. Opsahl-Ong, P. R. Granfors, I. Levis, C. J. D'Orsi, and R. E. Hendrick, "Full breast digital mammography with an amorphous silicon-based flat panel detector: Physical characteristics of a clinical prototype," Med. Phys. **27**, 558–567 (2000).

[48]K. W. Jee, L. E. Antonuk, Y. El-Mohri, and Q. H. Zhao, "System performance of a prototype flat-panel imager operated under mammographic conditions," Med. Phys. **30**, 1874–1890 (2003).

[49]B. Lazzari, G. Belli, C. Gori, and K. Nykanen, "Physical characteristics of a clinical prototype for full-field digital mammography with an *a*-Se flat panel detector," Proc. SPIE **5030**, 656–666 (2003).

Effect of Display Resolution
on the Detection of Mammographic Lesions

Robert S Saunders, Jr [a,b], Ehsan Samei [a,b,c], Jeffrey Johnson,[d] Jay Baker[a]

[a] Duke Advanced Imaging Labs, Department of Radiology, Duke University Medical Center
[b] Department of Physics, Duke University
[c] Department of Biomedical Engineering, Duke University
[d] Siemens Corporate Research

ABSTRACT

For diagnosis of breast cancer by mammography, the mammograms must be viewed by a radiologist. The purpose of this study was to determine the effect of display resolution on the specific clinical task of detection of breast lesions by a human observer. Using simulation techniques, this study proceeded through four stages. First, we inserted simulated masses and calcifications into raw digital mammograms. The resulting images were processed according to standard image processing techniques and appropriately windowed and leveled. The processed images were blurred according to MTFs measured from a clinical Cathode Ray Tube display. JNDMetrix, a Visual Discrimination Model, examined the images to estimate human detection. The model results suggested that detection of masses and calcifications decreased under standard CRT resolution. Future work will confirm these results with human observer studies. (This work was supported by grants NIH R21-CA95308 and USAMRMC W81XWH-04-1-0323.)

Keywords: Image Quality, Mammography, Simulation, Task-Based Assessment

1. INTRODUCTION

After a digital mammogram has been acquired, a human observer must view the data in order to detect or diagnose disease. The display device, therefore, assumes a crucial role in the imaging chain. While several researchers have given significant attention to the quality of image acquisition,[1-9] fewer investigators have measured the impact of display devices.[10-13] To understand this impact, studies must evaluate the physical properties of these devices. However, while physical characterization remains important, display quality must ultimately be described in terms of the clinical task in question.[14-16] This study considered this type of question, examining the impact of display resolution on the detection of mammographic lesions.

A Cathode Ray Tube (CRT) display serves as a common mammographic display device.[17] As a CRT ages, its resolution becomes progressively more degraded, leading to lower display quality over time.[18] The purpose of this study was to consider how this degradation in resolution impacted the clinical utility of a CRT display, specifically the detection of breast masses and calcifications.

2. METHODS AND MATERIALS

In this study, first simulated masses and calcifications were inserted into digital mammograms. We applied basic image processing techniques to these images and adjusted the window and level appropriately. Next, we blurred the images according to three different resolution settings measured from a CRT display. Finally, a model observer viewed each of these images to estimate the detection probabilities under each blur setting. The following describes the details of these steps.

2.1 Acquisition of Digital Mammographic Backgrounds

Digital mammographic images were acquired on a clinical flat-panel cesium iodide-based digital mammography system (Senographe 2000D, GE Medical Systems, Milwaukee, WI). Previous studies have characterized the

physical characteristics of this digital mammography system.[19, 20] Images used in this study were normal craniocaudul view mammograms acquired with a molybdenum anode with molybdenum or rhodium filtration. The beam energies for the images ranged from 25 to 30 kVp and compressed breast thicknesses extended from 2.7 cm to 7.3 cm with varying glandular and adipose tissue composition.

2.2 Lesion Simulation

Simulated breast lesions were placed in the center of mammographic images using an established procedure for simulating masses and calcifications with attributes similar to those of real mammographic lesions.[21, 22] Breast mass simulation proceeded through three stages, as illustrated in Figure 1. The first stage sets each pixel of an array to its equivalent major axis value,

$$b = \sqrt{[(y - y_0)Cos\,[\alpha] - (x - x_0)Sin\,[\alpha]]^2 + \frac{1}{c^2}[(x - x_0)Cos\,[\alpha] + (y - y_0)Sin\,[\alpha]]^2} \tag{1}$$

where (x_0, y_0) represent the center of the mass, α determines the angular orientation of the mass, and c corresponds to the ratio of the minor axis length to the major axis length. The second stage introduced non-uniformities in the mass border by multiplying the elliptical rings with a border deviation profile with a given variance and power spectrum. The final stage converted the equivalent major axis values to detector gray level values through the elliptical trace function.

The calcification procedure similarly required three stages. The first stage established the distribution of calcifications, using either a clustered or linear distribution. The second stage created individual calcification at each point specified by the calcification distribution through a series of morphological thickening and erosion operations. This resulted in a binary mask of the calcifications. The final stage added the binary mask to a background image with the appropriate contrast.

The spatial parameters for the simulation routine were determined from screen-film mammographic data obtained through the Digital Database of Screening Mammography.[23] These parameters remained applicable to digital mammographic backgrounds. However, the lesion contrast must be separately calculated for the digital case as the contrast in screen-film images were impacted by varying H&D characteristics. To determine the appropriate contrast for the simulated lesions, the xSpect x-ray simulation program[24] calculated the unit contrast for both masses and calcifications embedded in a 50% glandular/50% adipose breast imaged with a cesium-iodide detector. The contrasts were calculated for a molybdenum anode with molybdenum or rhodium filtration for each kVp and breast thickness. Contrast reduction by scattered radiation was also accounted for using previously published measurements.[25] The lesions were then inserted in mammographic backgrounds with the appropriate contrast and spatial features for the given anode, filtration, kVp, and breast thickness.
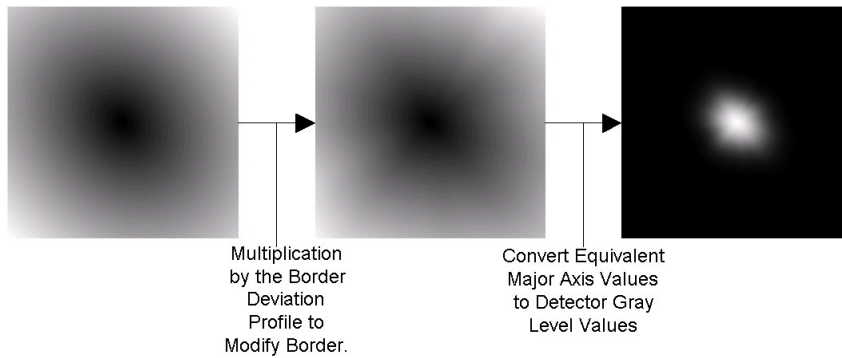

Multiplication by the Border Deviation Profile to Modify Border.

Convert Equivalent Major Axis Values to Detector Gray Level Values

FIG. 1 Schematic of mass simulation procedure. The three images illustrate the three steps in this system.
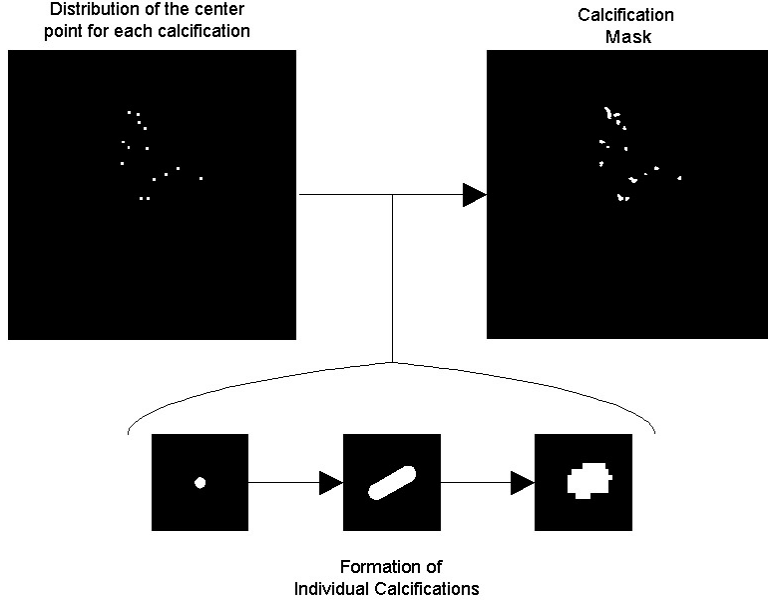
FIG. 2 Illustration of calcification simulation routine. This procedure considers each point in the distribution and creates a unique calcification on each point.

2.3 Image Processing

Most digital mammography systems employ post-processing algorithms to improve image display. A common technique separates the images into multiple frequency bands to improve contrast for specific frequencies. This study utilized a basic image processing algorithm that enhanced two frequency bands in the image.[26] The first stage augmented the higher frequency content of the image, while the second stage strengthened the content variations. The parameters for each stage were determined by visual analysis of the images. The first stage accentuated the sharp detail in the image through an unsharp masking procedure as,

$$I_{US} = I + SF(c) \cdot (I - \Sigma \otimes I) \tag{2}$$

where I represented the input image, Σ, the Gaussian kernel, had a standard deviation of 0.45 mm and width of 2.8 mm, and $SF(c)$, the sharpness factor, controlled the level of enhancement. To boost low contrast objects, a non-linear function was utilized for $SF(c)$ as,

$$SF(c) = \begin{cases} G \cdot \left(1 - \dfrac{c}{c_0}\right)^p + SF_0 & c \leq c_0 \\[2em] SF_0 & c > c_0 \end{cases} \tag{3}$$

with a gain, G, of 1, a contrast threshold, c_0, of 70, a contrast, c, equal to the absolute difference between the blurred image and original image, and a slope parameter, p, of 3. The second stage enhanced the mid-frequency components of the image, as

$$I_{Out} = (\xi \otimes I_{US}) + CF(c) \cdot (I_{US} - \xi \otimes I_{US}) \tag{4}$$

where ξ represented a Gaussian kernel with a standard deviation of 4.4 mm and $CF(c)$ controlled the level of contrast enhancement. The function $CF(c)$ had the same functional form as $SF(c)$, but utilized a gain, G, of 1.3.

Once the images were processed, observers window and level an image in order to produce an acceptable image appearance. To determine the window and level parameters for each mammogram, an experienced mammographer windowed and leveled each mammogram individually. A sigmoid transformation was fit to each window and level function, to provide a smooth transition at the extremes of the display pixel values. This transformation was represented as

$$I_{Final} = \alpha \left(1 - \frac{1}{1 + e^{-(Log[I_{Out}] - \delta)/\sigma}} \right) \tag{5}$$

where I_{Out} represented the processed mammogram, δ equaled the center of the sigmoid transition, and σ established the slope of the sigmoid transition.

2.4 Measurement of Display Characteristics

We measured the resolution properties of a five mega-pixel Cathode Ray Tube (CRT) display system (Barco MGD-521, p45 phosphor) with a 10-bit graphics controller (Barco 5MP1H). A charge-coupled device (CCD) camera (XCD-SX900, Sony Corporation, Tokyo, Japan) equipped with a macro lens (Rodgen 1:4, 28mm, Rodenstock, Munich, Germany) acquired images of line test patterns presented on the display. Two images from the recent TG18 test pattern set (TG18-RV50 and TG18-RH50) supplied a vertical and horizontal line, respectively.[27, 28] To remain in the quasi-linear range of the display, these patterns employed subtle lines, with 12% contrast from the background. We then computed the MTF from these line patterns using established methods. Full details of the measurement methodology has been reported in another publication.[29] We measured the MTF for the standard display resolution setting and two degraded resolution settings using the defocusing feature of the display. These measured MTFs are displayed in Figure 3.
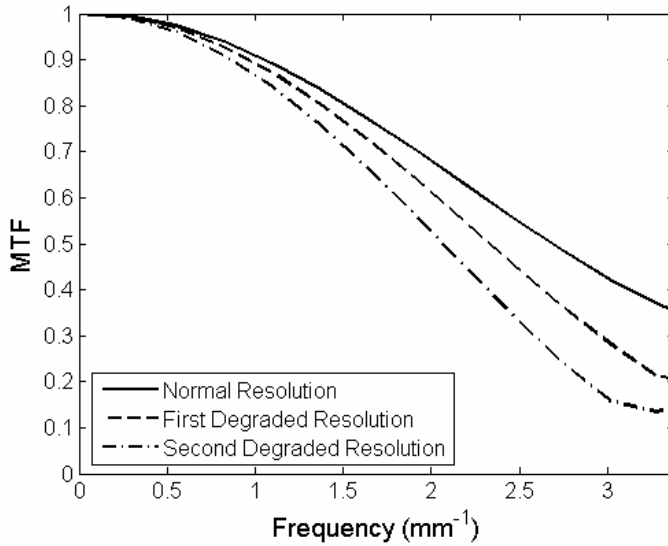

FIG. 3 Measured MTF for a CRT display under three different resolution settings.

2.5 Simulation of Image Display

A Resolution Modification routine, the details of which are disclosed in a previous publication,[30] simulated the blur effects of the CRT display. This routine altered the resolution of an input image according to an input MTF to produce a blurred version of the image. To accomplish this, the input mammogram was transformed to the frequency domain through an FFT. The frequency content of the image was then filtered by the display MTF. An inverse FFT transformed this modified frequency spectrum back to the spatial domain. This blurring was performed for each display resolution setting to produce multiple versions of each image.

2.6   Observer Model Experiment

A 5.12 cm x 5.12 cm region of interest (ROI) was extracted from the central breast area for analysis by a visual discrimination model (VDM). The Sarnoff JNDmetrix[33] VDM has been used to simulate the effects of display characteristics and image processing on the conspicuity of mammographic lesions.[11, 31, 32,34] For this study, the VDM compared a mammogram containing a lesion to the same mammogram without the lesion and computed a just-noticeable difference (JND) metric for the discriminability of those images by a human observer. The VDM first convolved the input images by an approximation of the point-spread function of the optics of the eye. The model simulated sampling of the image by retinal cones by performing a Gaussian convolution and then point sampling. Next, it computed the local contrast from the raw luminance image. The model applied a Laplacian pyramid to the data in order to isolate five frequency bands from the data. For each frequency band, the data was convolved with eight pairs of spatially oriented filters. The sensitivities and other parameters for these filters were determine by fitting model output to psychophysical data from sine-grating detection and discrimination experiments. The model squared each pair of filter output images and summed them to provide a phase-independent response. Next, the transducer stage derived the energy for each frequency band, normalizing this energy by the square of the appropriate grating contrast detection threshold. A sigmoid function was applied to each frequency bands to account for the visual contrast discrimination function.   The model incorporated the foveal sensitivity by averaging the outputs from the transducer step using a disk kernel. The final product of the model was a two-dimensional map of JND values, where each pixel indicated the discriminability of the two input images.

3.   RESULTS

Figure 4 illustrates the results when the VDM discriminates between mammographic images with simulated benign masses and those without simulated benign masses. The *perfect* resolution refers to images without any display blur, while the other three resolution settings refer to the measured MTFs in Figure 3. As expected, the model was better able to detect masses without any display blur. The difference between the three display blur settings remained much more modest.
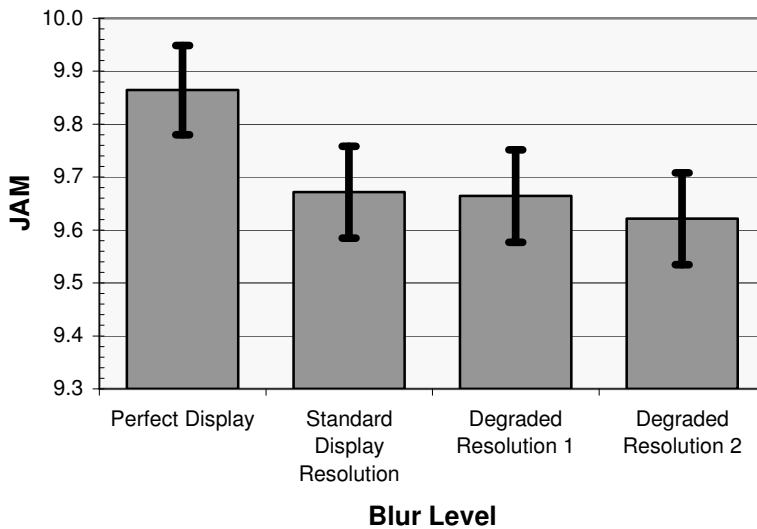

FIG. 4 JND Aggregate Measure (JAM) from the VDM comparing mammographic backgrounds with and without simulated benign masses for four different resolution settings. The benign masses had an average diameter of 3 mm. The error bars represent the 95% Confidence Interval.

Figure 5 illustrates the results when the VDM discriminates between mammographic images with simulated fine linear branching calcifications and those without simulated calcifications. The nomenclature in Figure 5 remains consistent with Figure 4. As expected, the model had a greater ability to detect these calcifications without any display blur. However, as the resolution of the CRT degrades, the detectability of calcifications decreased significantly. Similar model results were obtained for images with pleomorphic calcifications.
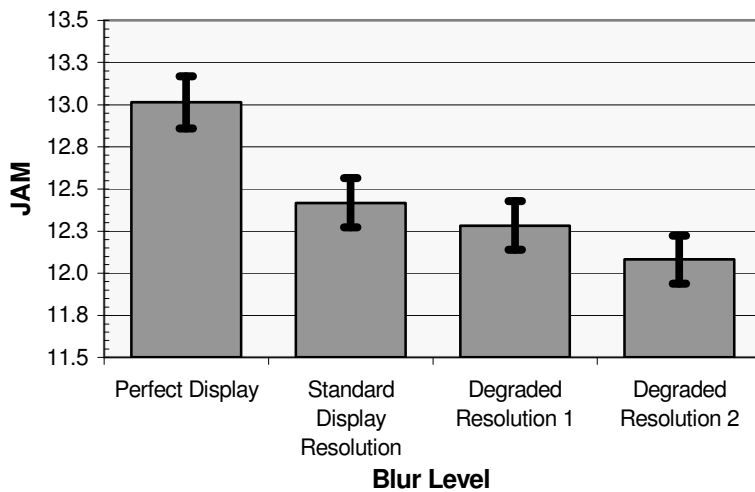
FIG. 5 JND Aggregate Measure (JAM) from the VDM comparing mammographic backgrounds with and without simulated fine linear branching calcifications for four different resolution settings.

4. DISCUSSION AND CONCLUSIONS

This study evaluated the impact of display blur on the detection of mammographic masses and calcifications. As an initial step, this study utilized a visual discrimination model to estimate detection by a human observer. These initial results suggested that detection of masses and calcifications decreased with standard CRT resolution. In addition, the model implies that the detection of calcifications, but not masses, declined as the resolution of the CRT degraded over time. This prediction seems reasonable because the conspicuity of small, fine structures in calcifications are more likely than larger objects, such as masses, to be affected by reductions in display resolution. The next phase of modeling will use VDM output to predict signal detectability within the framework of a channelized model observer. Future work must include human observer performance experiments to verify these estimates.

ACKNOWLEDGEMENTS

References

1   M. J. Flynn and E. Samei, "Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems," Med Phys **26**, 1612-1623 (1999).

2   W. Zhao, I. Blevis, S. Germann, J. A. Rowlands, D. Waechter and Z. S. Huang, "Digital radiology using active matrix readout of amorphous selenium: Construction and evaluation of a prototype real-time detector," Med Phys **24**, 1834-1843 (1997).

3   E. Samei and M. J. Flynn, "An experimental comparison of detector performance for computed radiography systems," Med Phys **29**, 447-459 (2002).

4   E. Samei and M. J. Flynn, "An experimental comparison of detector performance for direct and indirect digital radiography systems," Med Phys **30**, 608-622 (2003).

5   S. Vedantham, A. Karellas, S. Suryanarayanan, I. Levis, M. Sayag, R. Kleehammer, R. Heidsieck and C. J. D'Orsi, "Mammographic imaging with a small format CCD-based digital cassette: Physical characteristics of a clinical system," Med Phys **27**, 1832-1840 (2000).

[6]   O. Tousignant, Y. Demers, L. Laperriere, M. Nishiki, S. Nagai, T. Tomisaki, A. Takahashi and K. Aoki, "Clinical performance of a 14"" * 14"" real-time amorphous selenium flat-panel detector," Proc. SPIE **5030**, 71-6 (2003).

[7]   B. Lazzari, G. Belli, C. Gori and K. Nykanen, "Physical characteristics of a clinical prototype for full-field digital mammography with an a-Se flat panel detector," Proc. SPIE **5030**, 656-66 (2003).

[8]   S. S. Kang, J. K. Park, D. G. Lee, C. W. Mun, J. H. Kim and S. H. Nam, "X-ray detection properties of polycrystalline Cd/sub 1-x/Zn/sub x/Te detectors for digital radiography," Proc. SPIE **5030**, 853-60 (2003).

[9]   D. C. Hunt, O. Tousignant, Y. Demers, L. Laperriere and J. A. Rowlands, "Imaging performance of an amorphous selenium flat-panel detector for digital fluoroscopy," Proc. SPIE **5030**, 226-34 (2003).

[10] J. Fan, W. J. Dallas, H. Roehrig, E. A. Krupinski, K. Gandhi and M. K. Sundareshan, "Spatial noise of high-resolution liquid-crystal displays for medical imaging: quantitative analysis, estimation, and compensation," Proc. SPIE **5367**, 433-443 (2004).

[11] E.A. Krupinski, J. Johnson, H. Roehrig, J. Lubin. "Using a human visual system model to optimize soft-copy mammography display: influence of display phosphor", Acad Radiol **10**, 161-166 (2003).

[12] A. Badano, R. M. Gagne, R. J. Jennings, S. E. Drilling, B. R. Imhoff and E. Muka, "Noise in flat-panel displays with subpixel structure," Med Phys **31**, 715-723 (2004).

[13] H. Roehrig, J. Gaskill, J. Fan, A. Poolla and C. Martin, "In-field evaluation of the modulation transfer function of electronic display devices," Proc. SPIE **5367**, 456-463 (2004).

[14] A. R. Pineda and H. H. Barrett, "What does DQE say about lesion detectability in digital radiography?," Proc. SPIE **4320**, 561-9 (2001).

[15] H. H. Barrett, "Quality time with the fractious Fourier family [image quality]," Proc. SPIE **4392**, 9-21 (2001).

[16] H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," J Opt Soc Am A **7**, 1266-78 (1990).

[17] J. Fan, W. J. Dallas, H. Roehrig and E. A. Krupinski, "Improving visualization of digital mammograms on the CRT display system," Proc. SPIE **5029**, 746-753 (2003).

[18] K. D. Compton, "Factors affecting CRT display performance: specifying what works," Proc. SPIE **3976**, 412-23 (2000).

[19] S. Vedantham, A. Karellas, S. Suryanarayanan, D. Albagli, S. Han, E. J. Tkaczyk, C. E. Landberg, B. Opsahl-Ong, P. R. Granfors, I. Levis, C. J. D'Orsi and R. E. Hendrick, "Full breast digital mammography with an amorphous silicon-based flat panel detector: Physical characteristics of a clinical prototype," Med Phys **27**, 558-567 (2000).

[20] S. Suryanarayanan, A. Karellas and S. Vedantham, "Physical characteristics of a full-field digital mammography system," Nucl Instrum Methods **533**, 560-570 (2004).

[21] R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," Proc. SPIE **5372**, 242-250 (2004).

[22] R. S. Saunders Jr., E. Samei and J. A. Baker, "Simulation of Breast Lesions," 7th International Workshop on Digital Mammography (2004).

[23] M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer, "The Digital Database for Screening Mammography," in *Proceedings of the 5th International Workshop on Digital Mammography*, edited by (Medical Physics Publishing, Madison, WI, 2000), Vol. pp

[24] E. Samei, M. J. Flynn and W. R. Eyler, "Simulation of subtle lung nodules in projection chest radiography," Radiology **202**, 117-24 (1997).

[25] J. M. Boone, K. K. Lindfors, V. N. Cooper, 3rd and J. A. Seibert, "Scatter/primary in mammography: comprehensive results," Med Phys **27**, 2408-16 (2000).

[26] M. Stahl, T. Aach and S. Dippel, "Digital radiography enhancement by nonlinear multiscale processing," Med Phys **27**, 56-65 (2000).

[27] E. Samei, "TG18: A new development in display quality and performance evaluation," Med Phys **29**, 1314-1314 (2002).

[28] E. Samei and M. J. Flynn, "A method for in-field evaluation of the modulation transfer function of electronic display devices," Proc. SPIE **4319**, 599-607 (2001).

[29] R. S. Saunders, A. Farshchi and E. Samei, "Resolution and Noise Measurements of Selected Commercial Medical Displays," Med Phys **(In Review, January 2005)**,

30   R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med Phys **30**, 3006-3017 (2003).

31   W. B. Jackson, M. R. Said, D. A. Jared, J. O. Larimer, J. Gille and J. Lubin, "Evaluation of human vision models for predicting human observer performance," Proc. SPIE **3036**, 64-73 (1997).

32   E.A. Krupinski, J. Johnson, H. Roehrig, M. Engstrom, J. Fan, J. Nafziger, J. Lubin, W.J. Dallas. "Using a human visual system model to optimize soft-copy mammography display: influence of MTF compensation", Acad Radiol **10**, 1030-1035 (2003).

33   J. Lubin. "A Visual Discrimination Model for Imaging System Design and Evaluation." In Peli E (Ed.) *Visual Models for Target Detection and Recognition*. Singapore: World Scientific Publishers (1995).

34   E.A. Krupinski, J. Johnson, H. Roehrig, J. Nafziger, J. Fan, J. Lubin. "Use of a human visual system model to predict observer performance with CRT vs LCD display of images." J. Digital Imaging **17**, 258-263 (2004).

Simulation of Breast Lesions

Robert S Saunders, Jr [a,b], Ehsan Samei [a,b,c], Jay Baker[b]
[a] Department of Physics, Duke University
[b] Department of Radiology, Duke University Medical Center
[c] Department of Biomedical Engineering, Duke University

Abstract

Lesion simulation provides a tool when quantifying the utility of an imaging system in a detection task. For mammography, the important detection tasks are detecting breast masses and calcifications. In this study, we characterized the radiographic appearance of both masses and calcifications from images obtained from the Digital Database of Screening Mammography (DDSM). The characterization results were then used in a routine capable of creating simulated masses and calcifications. To verify the quality of this simulation routine, an observer performance experiment was conducted in which an observer was asked to discriminate between real and simulated lesions. The results were then analyzed using ROC analysis. The preliminary results showed an A_z of 0.59 for benign masses, 0.61 for malignant masses, and 0.58 for malignant calcifications. More observer studies are underway to enhance the statistical power of these results. (This work was supported by a grant from the NIH, R21-CA95308 and USAMRMC W81XWH-04-1-0323.)

1.   Introduction

A number of new full-field digital mammography systems with varying attributes have entered the clinical arena. It is important, therefore, to discover which systems are most appropriate for mammographic imaging. As the detection of breast cancer is the key task in mammography, a system should be judged in how well it aids in that task. Simulation techniques significantly facilitate such evaluations for a variety of detectors, breast densities, and lesion types.

One hurdle faced by mammography simulation is the lack of breast lesion models. For masses, previous work has used gaussian profiles, disks, and simulated lung nodule profiles. [1-4] For calcifications, the most common model has been to utilize masks extracted from real calcifications.[5, 6] This study adopted a different approach. First, we characterized the radiographic appearance of breast masses and calcifications from real mammograms. Then, we created simulated breast masses and calcifications emulating those characteristics. Our mass model was previously validated through a preliminary observer performance experiment.[7] This paper extends that work to microcalcifications.

2. Lesion Characterization

2.1 Breast Mass Characterization Procedure

Four categories of breast masses were chosen for characterization using the BI-RADS® lexicon.[8] Two types were typically benign, *oval circumscribed* and *oval obscured masses*, and two were typically malignant, *irregular ill-defined* and *irregular spiculated.* Sample mammograms containing these lesions were extracted from the University of South Florida's Digital Database for Screening Mammography (DDSM).[9] Characterization was performed on a 2.56 cm x 2.56 cm region of interest (ROI) surrounding the mass. All ROIs were converted to optical density values using the characteristic curve of the scanner.

The behavior of the masses was determined through a large-scale analysis and a small-scale analysis. The large-scale behavior was characterized through an elliptical trace, which examined the changes in optical density through concentric elliptical rings. The small-scale behavior was measured through a deviation profile that measured how the border of the lesion varied from an ellipse. These are shown graphically in figure 1.

Fig. 1. The elliptical trace, left, characterizes the large-scale behavior of the mass. The small-scale behavior is shown in the border deviation profile, right.

2.2 Breast Mass Characterization Results

Example characterization results for typically benign masses are shown in figure 2. The elliptical trace showed a sharp transition from the mass to the background, which was expected for a circumscribed border. The border deviation profile showed some deviations from the perfect ellipse, but the magnitude was fairly small. This was in contrast to the results for typically malignant masses, an example of which is shown in figure 3. The elliptical trace for these masses showed a very slow transition from the mass to the background. The border deviation profile illustrated strong deviations from the perfect ellipse. This was expected as the borders are ill-defined and the shape was irregular.
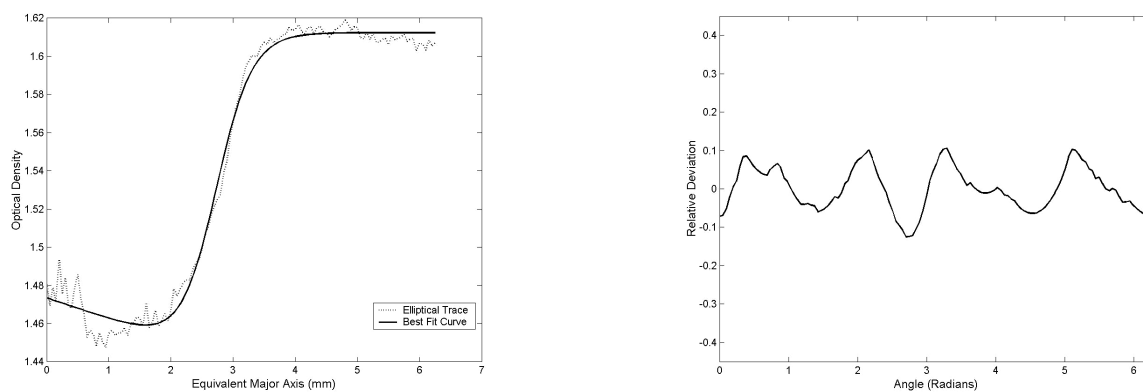

FIG. 2. Example characterization results for benign masses. The elliptical trace, left, shows a strong transition from mass to background while the border deviation profile, right, shows small deviations from the perfect ellipse.
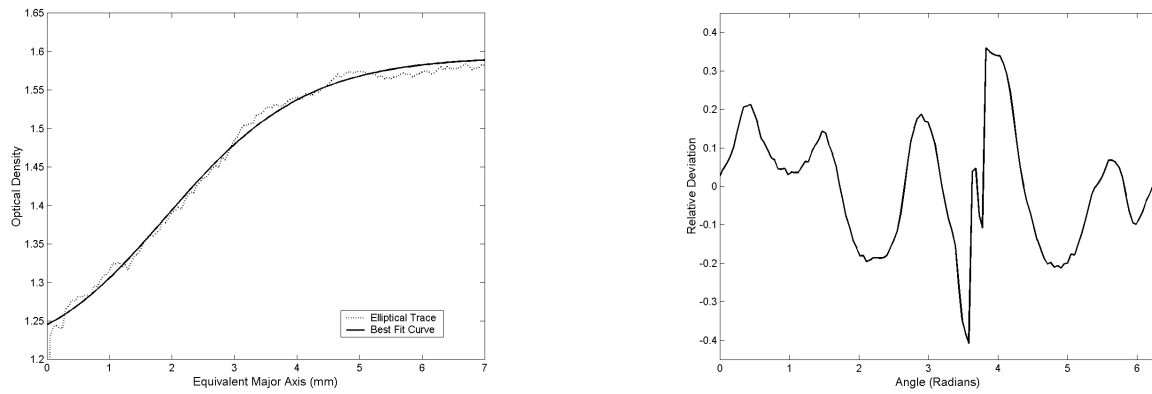
FIG. 3. Example characterization results for malignant masses. The elliptical trace, left, shows a smooth transition to background, and the border deviation profile, right, shows marked deviations from the perfect ellipse.

2.3   Calcification Characterization Procedure

Two categories of calcifications were chosen based on the BI-RADS® lexicon.[8] The two categories were *fine linear branching* and *pleomorphic*, referring to typically malignant lesions. The distribution studied for fine branching calcifications was linear, while the distribution studied for pleomorphic was clustered. Similar to masses, sample mammograms were drawn from the DDSM database.

To characterize the calcifications, a mask of the distribution was drawn. Measurements were then made on this binary mask. Three properties were measured for the individual calcifications: the major axis, minor axis, and the average contrast. Furthermore, the distributions for each calcification type were measured. For pleomorphic calcifications, the major axis and minor axis of the cluster were measured. For fine linear branching calcifications, the lengths of the lines were measured along with the angle between the lines of calcifications.

2.4   Calcification Characterization Results

The results from the calcification characterization are shown in table 1. The individual

calcifications results were similar for both pleomorphic and fine linear branching categories.

The distribution results established the mean shape for each distribution.

Table 1. Summary of calcification characterization results

Calcifications:	Pleomorphic	Fine Linear Branching
Major Axis (mm)	0.47 ± 0.11	0.43 ±0.13
Minor Axis (mm)	0.29 ±0.057	0.26 ± 0.045
Contrast	0.22 ± 0.13	0.34 ± 0.16
Distribution:		
Major Axis (mm)	8.0 ± 3.5	n/a
Minor Axis (mm)	7.1 ± 3.2	n/a
Line Length (mm)	n/a	6.2 ± 2.3
Angle (degrees)	n/a	50.8 ± 11.2

3.   Lesion Simulation

3.1   Mass Simulation

The mass simulation routine began with an array where each pixel was set equal to its equivalent

major axis value (given the eccentricity and center location). The border deviation effects were

then applied to this array. Finally, the array was transformed to optical density using the

elliptical trace profile. This is shown graphically in figure 4. Example masses are shown
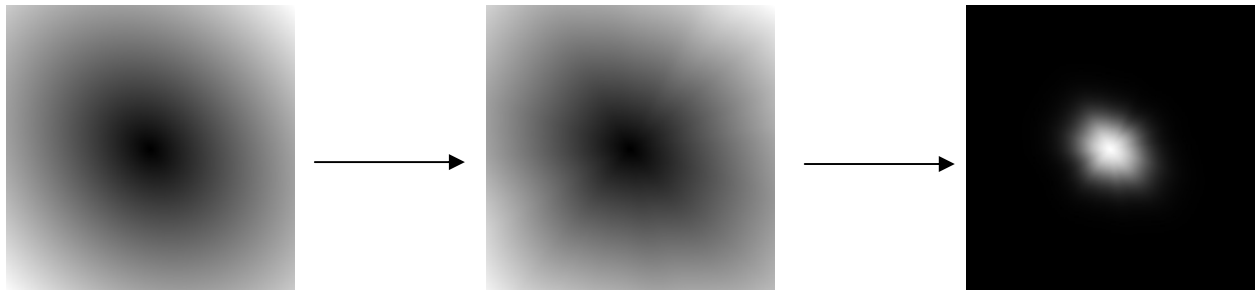
imbedded in backgrounds in figure 5.

FIG. 4. Graphical overview of mass simulation procedure. The image on left shows an array with pixel values equal to their equivalent major axis value. The border deviations are introduced in the center image. Finally, the image is transformed to optical density through the elliptical trace profile, which results in the final image on the right.

FIG. 5. Example simulated masses. The image on the left is a simulated benign mass, while the image on the right is a simulated malignant mass with an ill-defined border.

3.2   Calcification Simulation

The measured distribution results established a probability distribution for the individual calcification centers. For the pleomorphic category, the centers had a uniform probability density within an ellipse with a given major axis and minor axis length. For the fine, linear branching case, the centers had a uniform probability distribution along lines with a given mean length and relative angle between lines.

Given the desired number of individual calcifications, the simulation program sampled these distributions to determine the location of the centers of the individual calcifications. For each individual calcification, a line was drawn through this center at a random angle. The length of this line was equal to the major axis length of the individual calcifications. A morphological thickening operation was then applied, followed by a morphological eroding. These produced the shapes of the individual calcifications. The calcification distribution was then added to a normal background with a given contrast. Example simulated calcifications were shown in figure 6.
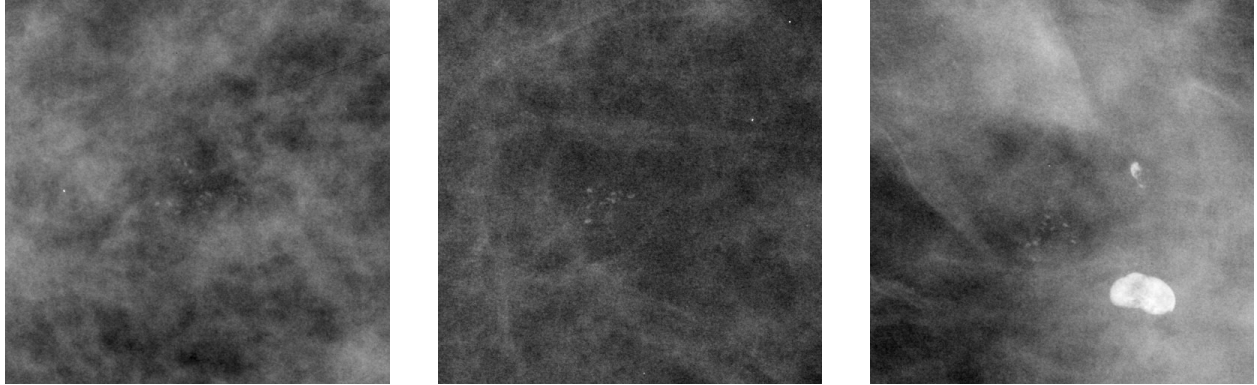
FIG. 6. Example simulated calcifications. The left image shows a simulated pleomorphic distribution, while the center and right image show simulated fine, linear branching calcifications.

4.   Observer Performance Experiment

4.1   Observer Protocol

To determine the quality of the simulation routines, an observer performance experiment was conducted. In this study, an experienced mammographer was asked to rate their confidence in whether a lesion was definitely real or definitely simulated. The simulation routine would be effective if a mammographer was unable to distinguish the difference between the simulated and real lesions. As this was a preliminary experiment, only one mammographer was used.

4.2   Observer Results

The histograms for the observer results for masses are shown in figure 7. In general, the distributions for real and simulated masses overlap considerably. The histogram for calcifications is shown in figure 8. Again, the histograms for real and simulated lesions overlap considerably.
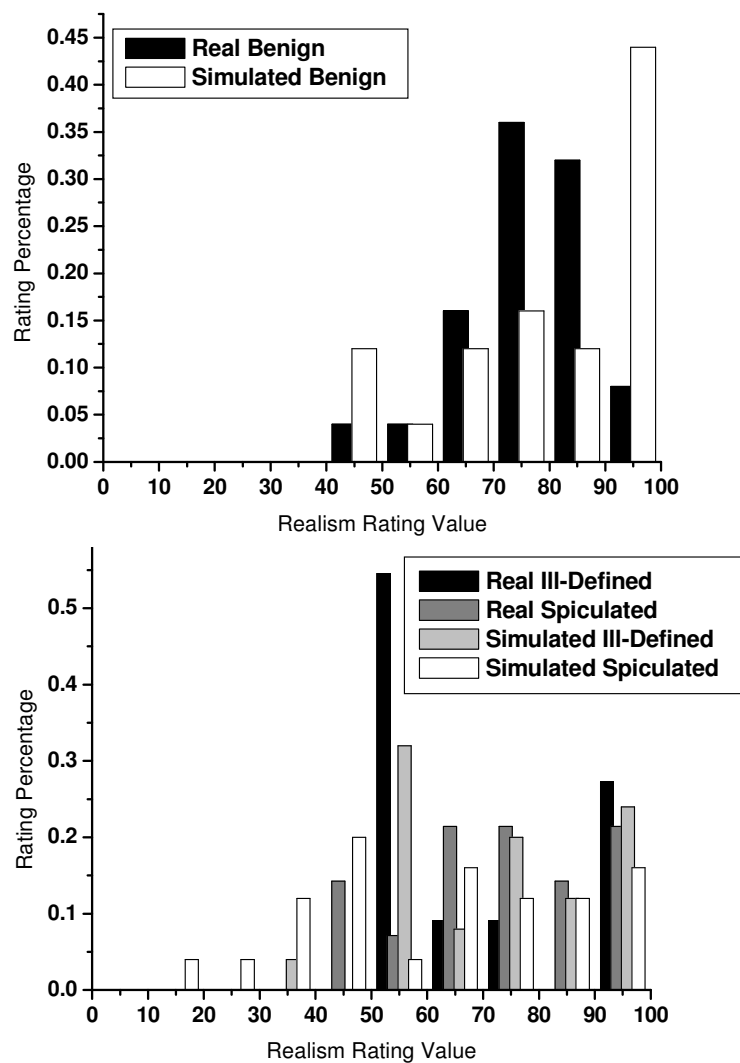
FIG. 7. Histograms of the rating frequency versus rating value for real and simulated masses. The results for typically benign masses are shown on the top while the typically malignant mass results are shown on the bottom. The typically malignant masses are further separated by border type for real and simulated masses.
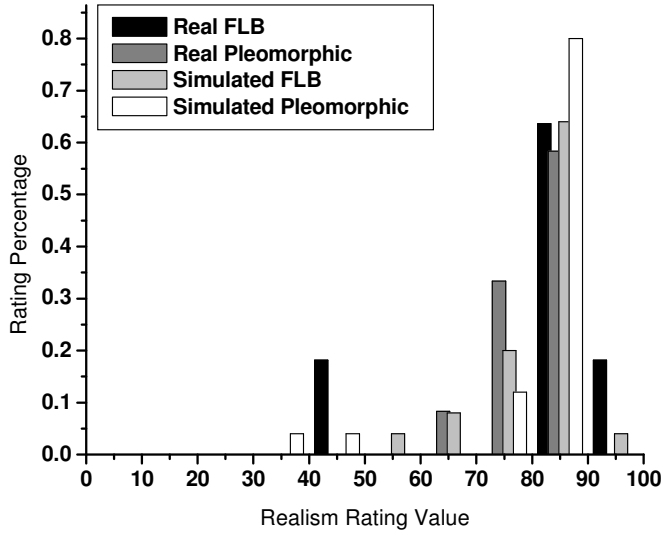
FIG. 8. Histogram of the rating frequency versus rating value for calcifications.

To quantify the degree of overlap, a Receiver Operating Characteristic Analysis was performed.[10] In this case, an A_z of 0.5 indicates that an observer was near chance in discriminating between real and simulated lesions. This analysis is summarized in table 2.

Table 2. Summary of ROC Analysis for discrimination between real and simulated lesions.

	A_z	σ
Benign Masses	0.59	0.08
Malignant Masses	0.61	0.07
Malignant Calcifications	0.58	0.07

5.   Conclusions

The characterization procedure undertaken in this study introduces a new way to describe breast lesions. The data from this characterization was then used in a new simulation routine that is capable of simulating breast masses and calcifications. Results from a preliminary observer performance experiment on these simulations indicate that our simulation routine produces high quality simulations of breast masses and calcifications. Further work is needed to validate the results of this preliminary observer performance experiment.

References

[1]   D. P. Chakraborty and H. L. Kundel, "Anomalous nodule visibility effects in mammographic images," Proc. SPIE **4324**, 68-76 (2001).

[2]   C. K. Abbey, M. P. Eckstein, S. S. Shimozaki, A. H. Baydush, D. M. Catarious and C. E. Floyd, "Human-observer templates for detection of a simulated lesion in mammographic images," Proc. SPIE **4686**, 25-36 (2002).

[3]   W. Huda, G. Qu, Z. Jing, B. G. Steinbach and J. C. Honeyman, "How does observer training affect imaging performance in digital mammography?," Proc. SPIE **3981**, 259-266 (2000).

[4]   A. E. Burgess, "Evaluation of detection model performance in power-law noise," Proc. SPIE **4324**, 123-32 (2001).

[5]   F. Lefebvre, H. Benali, R. Gilles and R. Di Paola, "A simulation model of clustered breast microcalcifications," Med Phys **21**, 1865-74 (1994).

[6]   M. Kallergi, M. A. Gavrielides, L. He, C. G. Berman, J. J. Kim and R. A. Clark, "Simulation model of mammographic calcifications based on the American College of Radiology Breast Imaging Reporting and Data System, or BIRADS," Acad Radiol **5**, 670-9 (1998).

[7]   R. S. Saunders and E. Samei, "Characterization of breast masses for simulation purposes," Proc. SPIE **5372**, 242-250 (2004).

[8]   C. D'Orsi, et al, Illustrated Breast Imaging Reporting and Data System (BI-RADS) (American College of Radiology, Reston, VA, 1998).

[9]   M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer, "The Digital Database for Screening Mammography," in *Proceedings of the 5th International Workshop on Digital Mammography*, edited by (Medical Physics Publishing, Madison, WI, 2000).

[10]   C. E. Metz, B. A. Herman and J. H. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat Med **17**, 1033-1053 (1998).

Physical Characterization of a Selenium-based
Full Field Digital Mammography Detector

Robert S. Saunders, Jr[a,b], Ehsan Samei[a,b,c], Joseph Lo[b,c], Jonathan Jesneck[b,c]

[a] Department of Physics, Duke University
[b] Duke Advanced Imaging Labs, Department of Radiology, Duke University Medical Center
[c] Department of Biomedical Engineering, Duke University

Abstract

Digital mammography has the potential to improve image quality for mammographic imaging. This study evaluated a selenium-based direct full-field digital mammographic imager (70 μm pixels) using a molybdenum anode operated at 28 kVp with inherent filtration of 30 μm molybdenum and an additional 2 mm of aluminum filtration. To capture the detector resolution, we measured the presampled modulation transfer function (MTF) using an edge method. The noise, summarized through the Normalized Noise Power Spectrum (NNPS), was measured by two-dimensional Fourier analysis of uniformly exposed radiographs. The detective quantum efficiency (DQE) was then computed from the measured MTF, NNPS, and ideal signal-to-noise ratio. For the Left-Right axis, the MTF reached the value of 0.2 and 0.1 at 12.7 mm^{-1} and 14.8 mm^{-1}, respectively. The DQE attained a maximum value of 53% at 1.45 mm^{-1} for the Left-Right axis. However, the DQE showed a strong dependence on exposure and frequency. The results indicated that this detector has high resolution, but it may be valuable to remove structured noise through improved calibration before clinical implementation. (The full data for this study are published as R.S. Saunders, Jr, E. Samei, J.L. Jesneck, and J.Y. Lo, "Physical characterization of a prototype selenium-based full field digital mammography detector," Med. Phys. 32(2) (2005).

1.   Introduction

The purpose of this work was to evaluate the physical characteristics of a selenium full-field digital mammography (FFDM) detector. Three different metrics of system performance were evaluated: the Modulation Transfer Function (MTF), Normalized Noise Power Spectrum (NNPS), and Detective Quantum Efficiency (DQE). As previous research has shown that selenium detectors can exhibit image lag and ghosting,[1] this research also examined the lag performance of the detector.

2.   Methods and Materials

2.1   Detector Description

The detector investigated in this study was a selenium-based flat-panel detector (Siemens Medical Systems, Erlangen, Germany). The detector was based on a 250 μm selenium photoconductive layer coupled to a storage capacitor and amorphous selenium switching transistor.[2] The active detector area was 23.296 cm x 28.672 cm consisting of 3328 x 4096 square pixels with a 70 μm pixel pitch. Prior to evaluation, the antiscatter grid supplied with the system was removed and gain and dead pixel corrections were performed according to manufacturer specifications. For most measurements, the standard detector cover was kept in place and the compression paddle was removed. For the MTF measurements, the detector cover was removed so that the edge device could be placed in contact with the detector.

2.2   Image Acquisition

The selenium detector was coupled to a high frequency multiphase x-ray generator (Mammomat Novation) for which the high-voltage accuracy was certified to be within ±5%. All images were

acquired with a large focal spot of 0.3 mm, nominal. We used the RQA-M2 technique, [3], which employed a molybdenum anode operated at 28 kVp, 30 μm molybdenum inherent filtration, and 2 mm aluminum added filtration. The image data were acquired in a raw format without any image post-processing applied. After acquisition, the images were transferred to a research computer as 14-bit, raw data for analysis.

For all image acquisitions, the exposure to the detector was measured free in air using a calibrated ionization chamber (1515 x-ray monitor with 10X5-6M dedicated mammography ionization chamber, Radcal Corporation, Monrovia, CA). The chamber was placed 17 cm above the detector to minimize contributions from backscatter. The exposure incident on the detector at 65 cm source to image distance (SID) was estimated from the measured exposure using the inverse-square law.

2.3   Linearity

The linearity of the detector was determined by exposing the detector to a wide range of uniform x-ray exposures for each of the radiographic techniques described above. The average pixel values were computed from a 14.3 x 14.3 cm region located near the chest wall section of the detector. From this, the relationships between mean pixel value and exposure were ascertained.

2.4   Modulation Transfer Function

The presampled MTF was measured using an edge method similar to that reported elsewhere.[4-9] A sharp edge test device, consisting of a polished 0.1 mm platinum-iridium edge, was placed in contact with the detector at 1 cm from the chest wall edge of the detector. The device was oriented with a 5–10 degree angle with respect to the pixel array. An image of the edge device

was then acquired using an exposure of 16.2 µC/kg (62.6 mR). The presampled modulation

transfer function (MTF) was then computed from the edge image using a method described in a

previous publication.[10] The MTF was computed along two orthogonal directions—the Chest

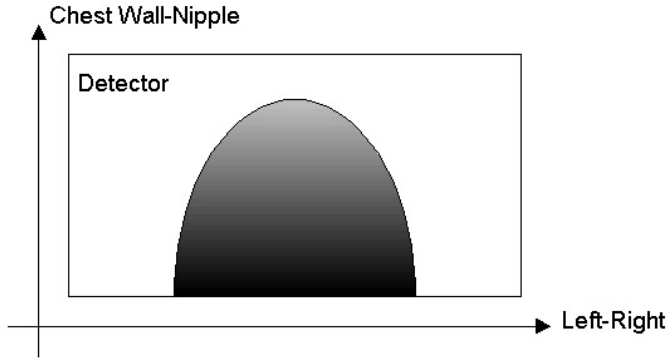Wall-Nipple (CN) axis and the Left-Right (LR), as shown in Figure 1.

FIG. 1. Coordinate system for measurements

2.5   Normalized Noise Power Spectrum

To characterize the system noise, flat-field images were acquired by exposing the detector to a

uniform x-ray beam. The exposure was simultaneously measured with the ionization chamber

reported above. The Normalized Noise Power Spectrum was then computed from the flat-field

images using previously published methods.[9, 10]

2.6   Detective Quantum Efficiency

The Detective Quantum Efficiency (DQE) was computed using the following equation:

$$DQE(u) = \frac{MTF^2(u)}{q_{Ideal}^2 \cdot E \cdot NNPS(u)} \tag{1}$$

where $MTF(u)$ represented the presampled modulation transfer function measured above, q_{Ideal} corresponded to the signal to noise (SNR) ratio per unit exposure for an ideal energy-integrating detector, E was the exposure value at the detector face at which the Normalized Noise Power Spectrum, $NNPS(u)$, was measured.[11, 12] The q_{Ideal} was computed with an x-ray simulation program (xSpect, Henry Ford Health System) that utilized a semiempirical model to simulate the x-ray spectra[13] and attenuation effects.[5]

2.7   Image Lag Measurement

The magnitude of multiplicative lag was characterized using the procedure described in IEC 62220-1.[14] First, an image was acquired of a uniform radiation field. The second image was then acquired of an edge device. After a specified delay time Δt, a third image was acquired of a uniform radiation field. The image data were then examined for two regions within the images. The first ROI, ROI_1, was placed in an area of the images that did not contain the edge device in image 2. The second ROI, ROI_2, was placed in an area that was inside the region covered by the edge device in image 2.
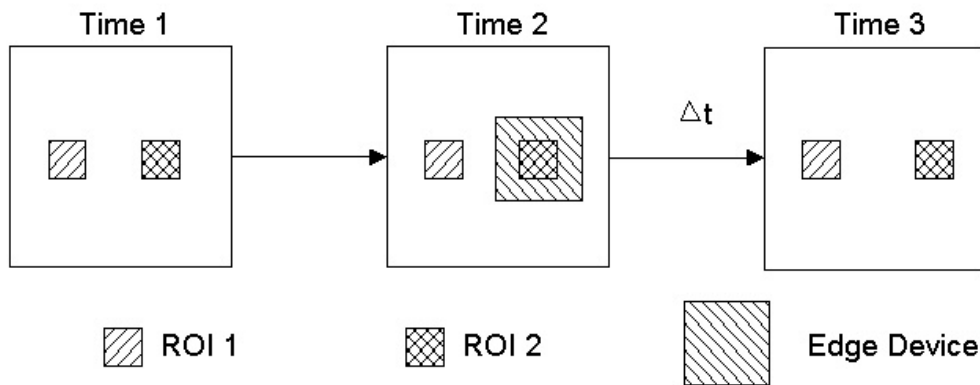

FIG. 2. Description of the lag measurement procedure.

The detector was judged to have acceptable lag effects for time delay Δt if it passed the following criterion[14]

$$\frac{\left|(\text{Image1}_{\text{ROI1}} - \text{Image1}_{\text{ROI2}}) - (\text{Image3}_{\text{ROI1}} - \text{Image3}_{\text{ROI2}})\right|}{\dfrac{\text{Image1}_{\text{ROI2}} + \text{Image3}_{\text{ROI2}}}{2}} \leq 0.005 \qquad (2)$$
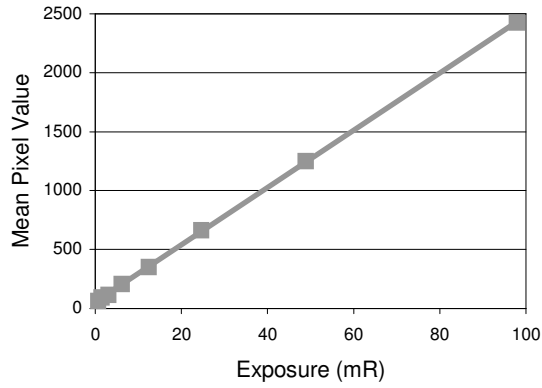
3. Results

FIG. 3. Plot of mean pixel value versus exposure. The system showed a very linear response with $r^2 > 0.999$.

The large area transfer characteristics of the detector are shown in figure 3. The detector maintains its linearity over two orders of magnitude in exposure. The MTF is shown in figure 4 for the CN and LR axis. The MTF along these axes diverge at higher spatial frequencies. The MTF curves are summarized in Table I for each axis.
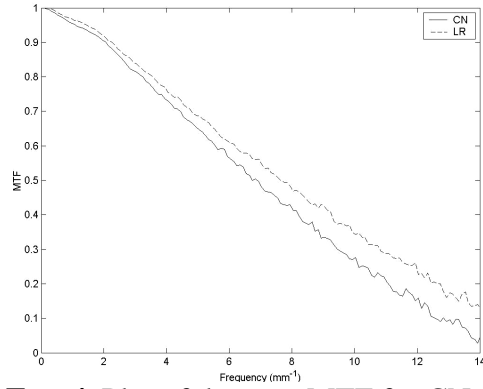
FIG. 4. Plot of detector MTF for CN and LR axes

	RQA-M2 (CN Axis)	RQA-M2 (LR Axis)
0.2 MTF	11.1 mm^{-1}	12.7 mm^{-1}
0.1 MTF	12.8 mm^{-1}	14.8 mm^{-1}
0.5 mm^{-1}	0.983	0.991
2.5 mm^{-1}	0.858	0.877
5.0 mm^{-1}	0.65	0.689

Table I. Summary of the detector's MTF properties

The radial traces of the NNPS multiplied by exposure are shown in figure 5 for each radiographic technique. The product of NNPS and exposure should remain constant for strictly quantum noise-limited detectors. However, the system exhibited exposure dependencies. For lower exposures, the magnitude of this metric first decreased and then increased with increasing exposure.
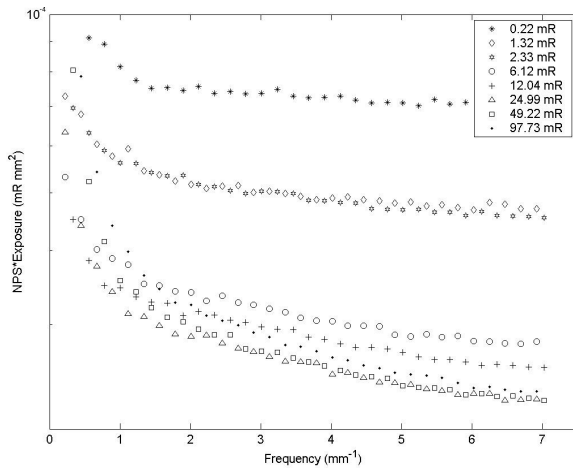

FIG. 5. Radial trace of NNPS for various exposure levels

Figure 6 shows the measured DQE. The DQE curves showed a decline at low frequency, which was expected from the strong low-frequency component of the NNPS. As well, the DQE increased with exposure for lower exposure values, reached a peak value, and then decreased for

higher exposures. This was also expected from the behavior of the NNPS with exposure. The DQE is summarized in table II.
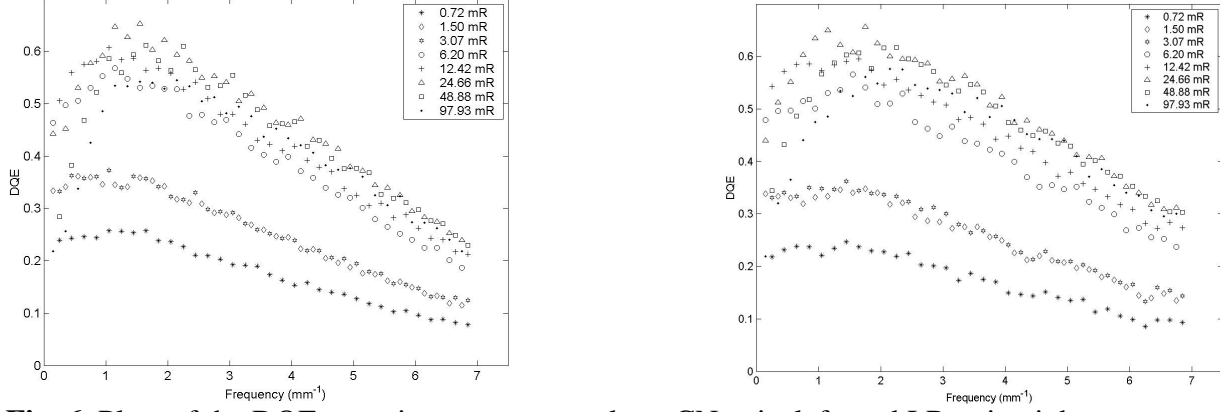

Fig. 6 Plots of the DQE at various exposures along CN axis, left, and LR axis, right.

Table II. Detector DQE properties for CN and LR axes at 1.6 µC/kg (6.2 mR)

	CN Axis	LR Axis
0.15 mm^{-1}	46%	47%
2.5 mm^{-1}	49%	50%
5.0 mm^{-1}	31%	34%
Peak	55% 1.25 mm^{-1}	53% 1.45 mm^{-1}

The results from lag measurements are summarized in table III. The images were acquired in the order indicated in table 3, with shorter delay time tests preceding longer delay time tests. In general, the image lag for the detector passed the test established by the IEC (Eq 2). However, an interesting phenomenon occurred for the 75 µGy exposure with 5 minute delay. A 200 µGy exposure was acquired 10 minutes before this exposure. It appeared that this high exposure still affected the detector after 10 minutes, as a 75 µGy exposure should not have caused a larger lag contribution after a 5 minute decay time than it would after a 3 minute decay time.

Table III. Summary of Multiplicative Lag Measurements

Exposure (µGy)	Decay Time (min)	Metric	Acceptable?
75	3	0.002	Yes
150	3	0.0048	Yes
200	3	0.0044	Yes
Ten Minute Wait			
75	5	0.047	No
150	5	0.013	No
200	5	0.0022	Yes

4.   Discussion

This prototype detector has excellent resolution properties, as shown by its MTF. There

appeared to be an asymmetry in the MTF, as it diverged for the CN and LR axes. As the edge

device was placed directly on the detector surface, it appeared unlikely that the focal spot caused

such asymmetries. Future work is needed to understand the cause of this asymmetry. The

prototype showed structured noise contributions, which led to a strong low-frequency

contribution to the NNPS. This structured noise also affected the DQE, in that the DQE curves

had a peak and then decreased for lower frequencies. Finally, image lag appeared to be within

the parameters established by IEC 62220-1,[14] but high exposures led to unusual behavior in

signal retention, even after a long decay. This prototype showed excellent promise and it is

expected that future work will correct the observed structured noise and lag phenomena with a

more robust calibration technique.

References

[1]   W. Zhao, G. DeCrescenzo and J. A. Rowlands, "Investigation of lag and ghosting in amorphous selenium flat-panel X-ray detectors," Proc. SPIE **4682**, 9-20 (2002).

[2]   J. G. Yorker, L. S. Jeromin, D. L. Y. Lee, E. F. Palecki, K. P. Golden and Z. Jing, "Characterization of a full-field digital mammography detector based on direct X-ray conversion in selenium," Proc. SPIE **4682**, 21-9 (2002).

[3]   International Electrotechnical Comisssion, Medical diagnostic X-ray equipment - Radiation conditions for use in the determination of characteristics (IEC-61267, Geneva, Switzerland, 2003).

[4]   E. Samei, M. J. Flynn, H. G. Chotas and J. T. Dobbins, III, "DQE of direct and indirect digital radiographic systems," Proc. SPIE **4320**, 189-97 (2001).

[5]   E. Samei and M. J. Flynn, "An experimental comparison of detector performance for computed radiography systems," Med Phys **29**, 447-459 (2002).

[6]   E. Samei and M. J. Flynn, "An experimental comparison of detector performance for direct and indirect digital radiography systems," Med Phys **30**, 608-622 (2003).

[7]   E. Samei, "Image quality in two phosphor-based flat panel digital radiographic detectors," Med Phys **30**, 1747-57 (2003).

[8]   E. Samei, M. J. Flynn and D. A. Reimann, "A method for measuring the presampled MTF of digital radiographic systems using an edge test device," Med Phys **25**, 102-113 (1998).

[9]   M. J. Flynn and E. Samei, "Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems," Med Phys **26**, 1612-1623 (1999).

[10]   R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," Med Phys **30**, 3006-3017 (2003).

[11]   J. C. Dainty and R. Shaw, Image science : principles, analysis and evaluation of photographic-type imaging processes (Academic Press, London ; New York, 1974).

[12]   J. Dobbins, "Image Quality Metrics for Digital Systems," in *Handbook of Medical Imaging*, edited by H. K. J. Beutel and R. V. Metter (SPIE, Washington, DC, 2000), Vol. 1, pp 163-222.

[13]   E. Storm, "Calculated bremsstrahlung spectra from thick tungsten targets," Phys. Rev. A **5**, 2328-38 (1972).

[14]   International Electrotechnical Commisssion, Medical electrical equipment - Characteristics of digital X-ray imaging devices - Part 1: Determination of the detective quantum efficiency (IEC 62220-1, Geneva, Switzerland, 2003).